TFW

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/611,363	07/01/2003	John R. Desjarlais	A-71486-2	4995

7590 09/22/2006

Robin M. Silva
DORSEY & WHITNEY LLP
Suite 3400
Four Embarcadero Center
San Francisco, CA 94111-4187

EXAMINER	
DEBERRY, REGINA M	

ART UNIT	PAPER NUMBER
1647	

DATE MAILED: 09/22/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/611,363	DESJARLAIS ET AL	
	Examiner	Art Unit	
	Regina M. DeBerry	1647	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *05 July 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-33* is/are pending in the application.

4a) Of the above claim(s) *2,4,12,14,17,19,22,24,27,28,31 and 33* is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,3,5-11,13,15,16,18,20,21,23,25,26,29,30 and 32* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *01 July 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *8/04,6/06*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

Status of Application, Amendments and/or Claims

Applicant's election without traverse of Group I (claims 1-32) and election of modifications at positions 223, 225, 226, 237 and 269 in the reply filed 05 July 2006 is acknowledged. In addition, all other claims reciting elected modifications (and combinations of those elected modifications) will be considered.

Claims 2, 4, 12, 14, 17, 19, 22, 24, 27, 28, 31 and 33 are withdrawn from further consideration pursuant to 37 CFR 1.142(b) as being drawn to a nonelected Group (or elected modification), there being no allowable generic or linking claim. Election was made **without** traverse in the reply filed on 17 April 2006.

Claims 1, 3, 5-11, 13, 15, 16, 18, 20, 21, 23, 25, 26, 29, 30 and 32 are under examination.

Inventorship

In view of the papers filed 16 April 2004, it has been found that this nonprovisional application, as filed, through error and without deceptive intent, improperly set forth the inventorship, and accordingly, this application has been corrected in compliance with 37 CFR 1.48(a). The inventorship of this application has been changed by the addition of Shannon A. Marshall.

The application will be forwarded to the Office of Initial Patent Examination (OIPE) for issuance of a corrected filing receipt, and correction of Office records to reflect the inventorship as corrected.

Information Disclosure Statement

The information disclosure statement(s)(IDS) filed 16 August 2004 and 30 June 2006 were received and comply with the provisions of 37 CFR §§1.97 and 1.98. They have been placed in the application file and the information referred to therein has been considered as to the merits.

Sequence Rules

The specification is not in compliance with 37 CFR 1.821-1.825 of the Sequence Rules and Regulations. When the description of a patent application discusses a sequence listing that is set forth in the "Sequence Listing" in accordance with paragraph (c) of the Sequence Rules and Regulations, reference must be made to the sequence by use of the assigned identifier (SEQ ID NO:), in the text and claims of the patent application.

37 CFR 1.821(a) presents a definition for nucleotide and/or amino acid sequences. This definition sets forth limits in terms of numbers of amino acids and/or numbers of nucleotides, at or above which compliance with the sequence rules is required. Nucleotide and/or amino acid sequences as used in 37 CFR 1.821 through 1.825 are interpreted to mean an unbranched sequence of four or more amino acids or an unbranched sequence of ten or more nucleotides. Please see MPEP section 2422.01.

The specification refers to sequences in Figure 2A, but does not identify the sequences by their sequence identifiers. Sequences appearing in drawings should be referenced in the corresponding Brief Description thereof. See 37 C.F.R. §1.58(a) and §1.83. Appropriate correction is required.

Applicant must submit a response to this Office Action and compliance with the sequence rules within the statutory period set for response to this Office Action.

Claim Rejections - 35 U.S.C. § 112

The following is a quotation of the first paragraph of 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

Claims 1, 3, 5-11, 13, 15, 16, 18, 20, 21, 23, 25, 26, 29, 30 and 32 are rejected under 35 U.S.C. 112, first paragraph, because the specification, while being enabling for:

a variant RANKL protein wherein said variant RANKL comprises modifications at positions R223M, R223E, R223Q, H225T, H225N, H225E, H225R, E226Q, E226D, E226R, Q237T, Q237K, Q237E, E269R, E269T, E269Q and E269K *in combination with mutation C221S/I247E,*

does not reasonably provide enablement for:

a variant RANKL protein wherein said variant RANKL comprises modifications at positions R223, H225, E226, Q237, E269.

The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention commensurate in scope with these claims.

The instant specification teaches that normal bone remodeling is a process in which new bone deposition by osteoblast is balanced through bone resorption by osteoclast. RANK is activated by the binding of its ligand, RANKL, which leads to differentiation, survival and fusion of pre-osteoclasts to form active bone resorbing osteoclast (page 1). The specification teaches that the present invention is directed at generating novel variants of human RANKL protein, comprising the extracellular domains of RANKL, which behave as RANKL antagonists or superagonists and modifications that confer soluble expression in E. coli.

The specification states that it has been observed that human RANKL forms inclusion bodies when expressed in E. coli. Soluble expression allows for efficient and cost-effective production and manufacturing of human RANKL variants (page 3, line 25-page 4, line 5). Thus, it is important that variant RANKL proteins are soluble. The specification teaches that only specific RANKL variants (C221S/I247A, C221S/I247D, C221S/I247K, C221S/I247Q and C221S/I247E) showed soluble expression in bacteria (page 39, Table 1). The specification teaches the construction of a RANKL variant library, comprising the solubility-imparting modification C221/I247E. The specification teaches the classification of RANKL variants. Specific RANKL variants exhibited non-

agonistic activity (i.e. inhibition of osteoclastogenesis with or without RANK receptor binding and/or inhibition of OPG binding)(Table 2 and pages 44-47).

It is known to those skilled in the art that certain positions in a sequence are critical to the protein's structure/function relationship, e.g. such as various sites or regions directly involved in binding, activity and in providing the correct three-dimensional spatial orientation of binding and active sites. These or other regions may also be critical determinants of antigenicity. These regions can tolerate only relatively conservative substitutions or no substitutions (see Wells, 1990, Biochemistry 29:8509-8517; Ngo *et al.,* 1994, The Protein Folding Problem and Tertiary Structure Prediction, pp. 433-440 and 492-495). For instance, a point mutation change from glutamine to aspartic acid at position 226 enables a variant RANKL protein to go from non-binding to binding of RANK receptor. A point mutation change from glutamine to threonine at position 269 still enables a variant RANKL protein to bind RANK receptor but inhibits the variant from binding OPG. It would be apparent to one skill in the art, that the effects of these types of changes are largely unpredictable as to which ones have a significant effect versus not. The instant specification teaches specific variant RANKL proteins, which are suitable. Therefore, the recitation of any RANKL variant protein results in an unpredictable and therefore unreliable correspondence between the claimed biomolecule and the indicated similar biomolecule of known function and therefore lacks support regarding enablement.

Lastly, claim 32 recites, "a pharmaceutical composition comprising a variant RANKL protein according to claim 1 and a pharmaceutical carrier" and thus reads on

use of that composition for treatment/therapy. The specification fails to disclose a direct correlation (working examples, animal models, etc.) between the use of the instant invention and a method for treatment in subjects. Specific RANKL variants were able to inhibit osteoclastogenesis *in vitro*. This activity is not predictive of the activity RANKL variants might have *in vivo*. Freshney (Culture of Animal Cells, A Manual of Basic Technique, Alan R. Liss, Inc., 1983, New York, p4) teaches that it is recognized in the art that there are many differences between cultured cells and their counterparts *in vivo*. These differences stem from the dissociation of cells from a three-dimensional geometry and their propagation on a two-dimensional substrate. Specific cell interactions characteristic of histology of the tissue are lost. The culture environment lacks the input of the nervous and endocrine systems involved in homeostatic regulation *in vivo*. Without this control, cellular metabolism may be more constant *in vitro* but may not be truly representative of the tissue from which the cells were derived. This has often led to tissue culture being regarded in a rather skeptical light. Thus, it could not be predicted that the cell culture data presented in the specification would be in any way correlative with therapeutic agents for *in vivo* treatments.

Due to the large quantity of experimentation necessary to generate the infinite number of derivatives recited in the claims and screen same for activity and the large quantity of experimentation necessary to show a correlation between a pharmaceutical composition comprising a RANKL variant and treatment of a specific disease/condition (including amounts and routes of administration for treatment in mammals), the absence of working examples directed to same, the complex nature of the invention, the state of

the prior art which establishes the unpredictability of the effects of mutation on protein structure and function, and the breadth of the claims which fail to recite any functional limitations, undue experimentation would be required of the skilled artisan to make and/or use the claimed invention in its full scope.

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 1, 3, 5-11, 13, 15, 16, 18, 20, 21, 23, 25, 26, 29, 30 and 32 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The instant claims are indefinite in the recitation of amino acid positions in the absence of a referenced SEQ ID NO:. It is not clear what sequence is intended by the claims and thus the metes and bounds of the claims cannot be determined by one skilled in the art.

Claim Objections

Claim 3 is objected to because of the following informalities: The word "substitution" is misspelled. Appropriate correction is required.

Conclusion

No claims are allowed.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Regina M. DeBerry whose telephone number is (571) 272-0882. The examiner can normally be reached on 9:00 a.m.-6:30 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Brenda G. Brumback can be reached on (571) 272-0961. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

RMD
9/13/06

MARIANNE P. ALLEN
PRIMARY EXAMINER
AU 1647

9/18/2006

Substitute PTO/SB/08A (07-05)
Approved for use through 07/31/2006. OMB 0651-0031
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Substitute for form 1449A/PTO
(Modified)

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(use as many sheets as necessary)

Sheet 1 of 1

Complete if Known	
Application Number	10/611,363
Filing Date	July 1, 2003
First Named Inventor	DESJARLAIS, John R.
Art Unit	1647
Examiner Name	~~FBA~~ R.M. Debary
Attorney Docket Number	Docket A-71486-2 (463077-240)

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	A1 ††	US-2002-061525 A1	05-23-2002	Rodrigo et al.	
	A2 †	US-6242213 B1	06-05-2001	Anderson	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3] Number[4] Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear	T[6]
	B1 †††	WO 00/15807 A *	03-23-2000	M & E Biotech		
	B2 ††	WO 00/67034 A	11-09-2000	Immunex Corporation		
	B3 †††	WO 01/25277 A	04-12-2001	Maxygen APS		
	B4 †††	WO 01/42298 A	04-21-2001	Kombluth		
	B5 †††	WO 01/64889 A	09-07-2001	Xencor		
	B6 ††	WO 02/18445 A	03-07-2002	Biogen, Inc.		
	B7 ††	WO 02/36141 A	05-10-2002	Immunex Corporation		
	B8 ††	WO 03/006154 A	01-23-2003	Xencor, Inc.		
	B9 ††	WO 03/029420 A	04-10-2003	Genentech, Inc.		
	B10 ††	WO 03/059281 A	07-24-2003	Xencor		
	B11 ††	WO 2004/081043 A	09-23-2004	Xencor		
	B12 ††	WO 2004/089982 A	10-21-2001	Xencor		
	B13 ††	WO 2005/035570 A	04-21-2005	Xencor, Inc.		
	B14 ††	WO 99/12965 A	03-18-1999	Biogen, Inc.		

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[6]
	C1 †††	Lacey D. L. et al. "Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation" *Cell* Cell Press, Cambridge, MA, US, vol. 93, April 17, 1998, pps. 165-176 *	
	C2 †††	Steed, P.M. et al. "Inactivation of TNF signaling by rationally designed dominant-negative TNF variants" *Science (American Assoc for the Advancement of Science)* vol. 301, no. 5641, (2003) pps. 1895-1898	
	C3 †††	Yasuda H. Et. Al. "Osteoclast Differentiation factor is a ligand for osteoprotegerin/osteoclastogeneis –Inhibitory factor and is identical to TRANCE/RANKYL" *Proc. Of Natl. Acad of Sci.* , (1998) vol. 95, pps. 3597-3602 *	

4818-1235-2513\16/26/2006/10:32 AM

Examiner Signature	R.M. Debary	Date Considered	9/13/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. [1]Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Substitute for form 1449A/PTO
(Modified)

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(use as many sheets as necessary)

Sheet 1 of 6

Complete if Known	
Application Number	10/611,363
Filing Date	July 1, 2003
First Named Inventor	DESJARLAIS, John R.
Art Unit	1647
Examiner Name	To Be Assigned R.M. DeSanto
Attorney Docket Number	A-71486-2

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	U.S. Patent Document Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	A1	5,843,678	12-01-1998	Boyle	
	A2	6,017,729	01-25-2000	Anderson et al.	
	A3 *	6,188,965 B1	02-13-2001	Mayo et al.	
	A4 *	6,269,312 B1	07-13-2001	Mayo et al.	
	A5	6,316,408 B1	11-13-2001	Boyle	
	A6 *	6,403,312 B1	06-11-2002	Dahiyat et al.	
	A7 *	6,708,120 B1	03-16-2004	Mayo et al.	
	A8 *	2001/0032052 A1	10-18-2001	Mayo et al.	
	A9 *	2001/0039480 A1	11-08-2001	Mayo et al.	
	A10 *	2002/0004706 A1	01-10-2002	Mayo et al.	
	A11 *	2002/0009780 A1	01-24-2002	Dahiyat et al.	
	A12 *	2002/0048772 A1	04-25-2002	Dahiyat et al.	
	A13 *	2002/0090648 A1	07-11-2002	Dahiyat et al.	
	A14 *	2002/0106694 A1	08-08-2002	Mayo et al.	
	A15 *	2002/0110868 A1	08-15-2002	Dahiyat et al.	
	A16	2003/0013651 A1	01-16-2003	Lam et al.	
	A17 *	2003/0049654 A1	03-13-2003	Dahiyat et al.	
	A18 *	2003/0130827 A1	07-10-2003	Dahiyat et al.	
	A19 *	2003/0138401 A1	07-24-2003	Dahiyat et al.	
	A20 *	2003/0166559 A1	09-04-2003	Desjarlais et al.	
	A21 *	2003/0219864 A1	11-27-2003	Desjarlais et al.	
	A22 *	2004/0043429 A1	03-04-2004	Dahiyat et al.	
	A23 *	2004/0043430 A1	03-04-2004	Dahiyat et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[2] Number[3] Kind Code[4] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear	T[5]
	B1 *	EP 0 974 111 B1	01-26-2000	California Institute of Technology		
	B2 *	WO 98/47089 A1	10-22-1998	California Institute of Technology		
	B3	WO 99/29865 A3	06-17-1999	The Rockefeller University		
	B4 *	WO 00/15807 A1	03-23-2000	M & E Biotech		
	B5 *	WO 00/23564 A2	04-27-2000	Xencor, Inc.		

Examiner Signature	R.M. DeSanto	Date Considered	9/3/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.
[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Substitute for form 1449A/PTO (Modified)				Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				Application Number	10/611,383
				Filing Date	July 1, 2003
				First Named Inventor	DESJARLAIS, John R.
				Art Unit	1647
(use as many sheets as necessary)				Examiner Name	~~To Be Assigned~~ R.M. DeBerry
Sheet	2	of	6	Attorney Docket Number	A-71486-2

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3] Number[4] Kind Code[5] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear	T[6]
	B6 *	WO 01/59066 A2, A3	08-16-2001	Xencor, Inc.		
	B7 *	WO 03/014325 A2, A3	02-20-2003	Xencor		
	B8	WO 03/033663 A2	04-24-2003	Barnes-Jewish Hospital		
	B9	WO 03/033664 A2	04-24-2003	Barnes-Jewish Hospital		

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[6]
	C1	ADLER, S.H., and TURKA L.A., "Immunotherapy as a means to induce transplantation tolerance," Curr Opin Immunol, 2002 Oct, 14(5):660-5.	
	C2	ALATALO S.L., et al., "Rapid Screening Method for Osteoclast Differentiation in Vitro That Measures Tartrate-resistant Acid Phosphatase 5b Activity Secreted into the Culture Medium," Clin Chem, 2000 Nov, 46(11):1751-4.	
	C3	ANDERSON DM, et al., "A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function," Nature, 1997 Nov 13, 390(6656):175-9.	
	C4	ARRON JR, and CHOI Y, "Bone versus immune system," Nature, 2000 Nov 30, 408(6812):535-6.	
	C5	ATKINS GJ, et al., "Osteoprotegerin inhibits osteoclast formation and bone resorbing activity in giant cell tumors of bone," Bone, 2001 Apr, 28(4):370-7.	
	C6	BACHMANN MF, and KOPF M., "Balancing protective immunity and immunopathology," Curr Opin Immunol, 2002 Aug, 14(4):413-9.	
	C7	BIANCHI L, et al., "A cluster region of AP-1 responsive elements is required for transcriptional activity of mouse ODC gene by hepatocyte growth factor," Arch Biochem Biophys, 2002 May 1, 401(1):115-23.	
	C8	BODMER JL, et al., "The molecular architecture of the TNF superfamily." Trends Biochem Sci., 2002 Jan, 27(1):19-26.	
	C9	CHILDS LM, et al., "In Vivo RANK Signaling Blockade Using the Receptor Activator of NF-kappaB:Fc Effectively Prevents and Ameliorates Wear Debris-Induced Osteolysis Via Osteoclast Depletion Without Inhibiting Osteogenesis," J Bone Miner Res., 2002 Feb, 17(2):192-9.	
	C10	COLLIN-OSDOBY P, et al., "Receptor Activator of NF-kappa B and Osteoprotegerin Expression by Human Microvascular Endothelial Cells, Regulation by Inflammatory Cytokines, and Role in Human Osteoclastogenesis," J Biol Chem., 2001 Jun 8, 276(23):20659-72.	
	C11	COMPSTON JE., "Bone marrow and bone: a functional unit," J Endocrinol., 2002 Jun, 173(3):387-94.	
	C12	CUROTTO de LAFAILLE MA, and LAFAILLE JJ., "CD4(+) regulatory T cells in autoimmunity and allergy," Curr Opin Immunol., 2002 Dec, 14(6):771-8.	
	C13	DARNAY BG, et al., "Characterization of the intracellular domain of receptor activator of NF-kappaB (RANK). Interaction with tumor necrosis factor receptor-associated factors and activation of NF-kappab and c-Jun N-terminal kinase," J Biol Chem., 1998 Aug 7, 273(32):20551-5.	

Examiner Signature		Date Considered	9/5/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Substitute PTO/SB/08A-B (08-03)
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

Substitute for form 1449A/PTO (Modified)				Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				Application Number	10/611,363
				Filing Date	July 1, 2003
				First Named Inventor	DESJARLAIS, John R.
				Art Unit	1647
(use as many sheets as necessary)				Examiner Name	To Be Assigned L.M. DeBlon
Sheet	3	of	6	Attorney Docket Number	A-71486-2

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[6]
	C14	DEYAMA Y, et al., "Histamine stimulates production of osteoclast differentiation factor/receptor activator of nuclear factor-kappaB ligand by osteoblasts," Biochem Biophys Res Commun., 2002 Oct 25, 298(2):240-6.	
	C15	FATA JE, et al., "The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development," Cell., 2000 Sep 29,103(1):41-50.	
	C16	GAO YH, et al., "Potential role of cbfa1, an essential transcriptional factor for osteoblast differentiation, in osteoclastogenesis: regulation of mRNA expression of osteoclast differentiation factor (ODF)," Biochem Biophys Res Commun., 1998 Nov 27, 252(3):697-702.	
	C17	GOATER JJ, et al., "Efficacy of ex vivo OPG gene therapy in preventing wear debris induced osteolysis," J Orthop Res. 2002 Mar, 20(2):169-73.	
	C18	GOOD CR, et al., "Immunohistochemical study of receptor activator of nuclear factor kappa-B ligand (RANK-L) in human osteolytic bone tumors," J Surg Oncol., 2002 Mar, 79(3):174-9.	
	C19	GORI F, et al., "The expression of osteoprotegerin and RANK ligand and the support of osteoclast formation by stromal-osteoblast lineage cells is developmentally regulated," Endocrinology, 2000 Dec, 141(12):4768-76.	
	C20	GOWEN M et al., "Emerging therapies for osteoporosis," Exp. Opin. on Emerging Drugs, 2000, 5(1):1-43	
	C21	GRAVALLESE EM, et al., "The role of TNF-receptor family members and other TRAF-dependant receptors in bone resorption," Arthritis Res., 2001, 3(1):6-12.	
	C22	HOFBAUER LC, et al., "Interleukin-1beta and tumor necrosis factor-alpha, but not interleukin-6, stimulate osteoprotegerin ligand gene expression in human osteoblastic cells," Bone, 1999 Sep, 25(3):255-9.	
	C23	HOFBAUER LC, et al., "Receptor activator of nuclear factor-kappaB ligand and osteoprotegerin: potential implications for the pathogenesis and treatment of malignant bone diseases," Cancer, 2001 Aug 1, 92(3):460-70.	
	C24	HONORE P, et al., "Osteoprotegerin blocks bone cancer-induced skeletal destruction, skeletal pain and pain-related neurochemical reorganization of the spinal cord," Nat Med. 2000 May;6(5):521-8.	
	C25	HOTOKEZAKA H, et al., "U0126 and PD98059, specific inhibitors of MEK, accelerate differentiation of RAW264.7 cells into osteoclast-like cells," J Biol Chem., 2002 Dec 6, 277(49):47366-72.	
	C26	HSU H, et al., "Tumor necrosis factor receptor family member RANK mediates osteoclast differentiation and activation induced by osteoprotegerin ligand," Proc Natl Acad Sci U S A., 1999 Mar 30, 96(7):3540-5.	
	C27	HUBER DM, et al., "Androgens suppress osteoclast formation induced by RANKL and macrophage-colony stimulating factor," Endocrinology, 2001 Sep, 142(9):3800-8.	
	C28	IKEDA T, et al., "Determination of three isoforms of the receptor activator of nuclear factor-kappaB ligand and their differential expression in bone and thymus," Endocrinology, 2001 Apr, 142(4):1419-26.	
	C29	ITO S, et al., "Crystal structure of the extracellular domain of mouse RANK ligand at 2.2-Å resolution," J Biol Chem., 2002 Feb 22, 277(8):6631-6.	
	C30	KANEDA T, et al., "Endogenous production of TGF-beta is essential for osteoclastogenesis induced by a combination of receptor activator of NF-kappa B ligand and macrophage-colony-stimulating factor," J Immunol., 2000 Oct 15, 165(8):4254-63.	
	C31	KITAZAWA R, and KITAZAWA S., "Vitamin D_3 augments osteoclastogenesis via vitamin D-responsive element of mouse RANKL gene promoter," Biochem Biophys Res Commun, 2002 Jan 18, 290(2):650-5.	
	C32	KITAZAWA R, et al., "Promoter structure of mouse RANKL/TRANCE/OPGL/ODF gene," Biochim Biophys Acta., 1999 Apr 14, 1445(1):134-41.	
	C33	KONG YY, et al., "Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand," Nature, 1999 Nov 18, 402(6759):304-9.	

Examiner Signature		Date Considered	9/5/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.
[1]Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Substitute for form 1449A/PTO (Modified)	Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	Application Number	10/611,363
	Filing Date	July 1, 2003
	First Named Inventor	DESJARLAIS, John R.
	Art Unit	1647
(use as many sheets as necessary)	Examiner Name	~~To Be Assigned~~ L.M. Debony
Sheet 4 of 6	Attorney Docket Number	A-71486-2

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[6]
	C34	LACEY DL, et al., "Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation," Cell, 1998 Apr 17, 93(2):165-76.	
	C35	LAM J, et al., "TNF-alpha induces osteoclastogenesis by direct stimulation of macrophages exposed to permissive levels of RANK ligand," J Clin Invest., 2000 Dec, 106(12):1481-8.	
	C36	LAM J, et al., "Crystal structure of the TRANCE/RANKL cytokine reveals determinants of receptor-ligand specificity," J Clin Invest., 2001 Oct, 108(7):971-9.	
	C37	LARK MW, and JAMES IE, "Novel bone antiresorptive approaches," Curr Opin Pharmacol., 2002 Jun, 2(3):330-7.	
	C38	LEAN JM, et al., "Osteoclast lineage commitment of bone marrow precursors through expression of membrane-bound TRANCE," Bone, 2000 Jul, 27(1):29-40.	
	C39	LI J, et al., "RANK is the intrinsic hematopoietic cell surface receptor that controls osteoclastogenesis and regulation of bone mass and calcium metabolism," Proc Natl Acad Sci U S A, 2000 Feb 15, 97(4):1566-71.	
	C40	LIU Y, et al., "Crystal structure of sTALL-1 reveals a virus-like assembly of TNF family ligands," Cell, 2002 Feb 8, 108(3):383-94.	
	C41	LOCKSLEY RM, et al., "The TNF and TNF receptor superfamilies: integrating mammalian biology," Cell, 2001 Feb 23, 104(4):487-501.	
	C42	LUM L, et al., "Evidence for a role of a tumor necrosis factor-alpha (TNF-alpha)-converting enzyme-like protease in shedding of TRANCE, a TNF family member involved in osteoclastogenesis and dendritic cell survival," J Biol Chem., 1999 May 7, 274(19):13613-8.	
	C43	MARTIN TJ, and GILLESPIE MT, "Receptor activator of nuclear factor kappa B ligand (RANKL): another link between breast and bone," Trends Endocrinol Metab., 2001 Jan-Feb, 12(1):2-4.	
	C44	MATSUZAKI K, et al., "Osteoclast differentiation factor (ODF) induces osteoclast-like cell formation in human peripheral blood mononuclear cell cultures," Biochem Biophys Res Commun., 1998 May 8, 246(1):199-204.	
	C45	McHUGH KP, et al., "Mice lacking beta3 integrins are osteosclerotic because of dysfunctional osteoclasts," J Clin Invest., 2000 Feb, 105(4):433-40.	
	C46	MENAA C, et al., "Enhanced RANK ligand expression and responsivity of bone marrow cells in Paget's disease of bone," J Clin Invest., 2000 Jun, 105(12):1833-8.	
	C47	NAKAGAWA N, et al., "RANK is the essential signaling receptor for osteoclast differentiation factor in osteoclastogenesis," Biochem Biophys Res Commun., 1998 Dec 18, 253(2):395-400.	
	C48	NAKASHIMA T, et al., "Protein expression and functional difference of membrane-bound and soluble receptor activator of NF-kappaB ligand: modulation of the expression by osteotropic factors and cytokines," Biochem Biophys Res Commun., 2000 Sep 7, 275(3):768-75.	
	C49	OYAJOBI BO, et al., "Therapeutic efficacy of a soluble receptor activator of nuclear factor kappaB-IgG Fc fusion protein in suppressing bone resorption and hypercalcemia in a model of humoral hypercalcemia of malignancy," Cancer Res., 2001 Mar 15, 61(6):2572-8.	
	C50	PENNINGER, Josef, "Bonas, Lymphocytes, and Mammalian Evolution," Society for Biomolecular Screening 8th Annual Conference, The Hague, The Netherlands, Sept. 22-26, 2002 (Abstract).	
	C51	QUINN JM, et al., "A combination of osteoclast differentiation factor and macrophage-colony stimulating factor is sufficient for both human and mouse osteoclast formation in vitro," Endocrinology, 1998 Oct, 139(10):4424-7.	
	C52	QUINN JM, et al., "The generation of highly enriched osteoclast-lineage cell populations," Bone, 2002 Jan, 30(1):164-70.	

Examiner Signature	L.M. Debony	Date Considered	9/5/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1]Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Substitute for form 1449A/PTO (Modified)				Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				Application Number	10/611,363
				Filing Date	July 1, 2003
				First Named Inventor	DESJARLAIS, John R.
				Art Unit	1647
(use as many sheets as necessary)				Examiner Name	To Be Assigned
Sheet	5	of	6	Attorney Docket Number	A-71486-2

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[6]
	C53	RODAN GA, and Martin TJ, "Therapeutic approaches to bone diseases," Science, 2000 Sep 1, 289(5484):1508-14.	
	C54	ROMAS E, et al., "Involvement of receptor activator of NFkappaB ligand and tumor necrosis factor-alpha in bone destruction in rheumatoid arthritis," Bone, 2002 Feb, 30(2):340-6.	
	C55	SABOKBAR A, et al., "Two distinct cellular mechanisms of osteoclast formation and bone resorption in periprosthetic osteolysis," J Orthop Res., 2003 Jan, 21(1):73-80.	
	C56	SAIDENBERG KERMANAC'H N, et al., "Osteoprotegerin and inflammation," Eur Cytokine Netw., 2002 Apr-Jun, 13(2):144-53.	
	C57	SCHLONDORFF J, et al., "Biochemical and pharmacological criteria define two shedding activities for TRANCE/OPGL that are distinct from the tumor necrosis factor alpha convertase," J Biol Chem., 2001 May 4, 276(18):14665-74.	
	C58	SCHOPPET M, et al., "RANK ligand and osteoprotegerin: paracrine regulators of bone metabolism and vascular function," Arterioscler Thromb Vasc Biol., 2002 Apr 1, 22(4):549-53.	
	C59	SENIOR K., "Vaccinating against bone destruction," Drug Discov Today, 2001 Dec 15, 6(24):1243-1244.	
	C60	SHEVDE NK, et al., "Estrogens suppress RANK ligand-induced osteoclast differentiation via a stromal cell independent mechanism involving c-Jun repression," Proc Natl Acad Sci U S A, 2000 Jul 5, 97(14):7829-34.	
	C61	SRIVASTAVA S, et al., "Estrogen decreases osteoclast formation by down-regulating receptor activator of NF-kappa B ligand (RANKL)-induced JNK activation," J Biol Chem., 2001 Mar 23, 276(12):8836-40.	
	C62	SUDA T, et al., "Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families," Endocr Rev., 1999 Jun, 20(3):345-57.	
	C63	TAKAHASHI N, et al., "A new member of tumor necrosis factor ligand family, ODF/OPGL/TRANCE/RANKL, regulates osteoclast differentiation and function," Biochem Biophys Res Commun., 1999 Mar 24, 256(3):449-55.	
	C64	TAKAYANAGI H, et al., "RANKL maintains bone homeostasis through c-Fos-dependent induction of interferon-beta," Nature, 2002 Apr 18, 416(6882):744-9.	
	C65	TAKAYANAGI H, et al., "Signaling crosstalk between RANKL and interferons in osteoclast differentiation," Arthritis Res., 2002, 4(Suppl 3):S227-32.	
	C66	TEITELBAUM SL., "Bone resorption by osteoclasts," Science, 2000 Sep 1, 289(5484):1504-8.	
	C67	THEILL LE, et al., "RANK-L and RANK: T cells, bone loss, and mammalian evolution," Annu Rev Immunol., 2002, 20:795-823.	
	C68	TSUDA E, et al., "Isolation of a novel cytokine from human fibroblasts that specifically inhibits osteoclastogenesis," Biochem Biophys Res Commun., 1997 May 8, 234(1):137-42.	
	C69	UDAGAWA N, et al., "Origin of osteoclasts: mature monocytes and macrophages are capable of differentiating into osteoclasts under a suitable microenvironment prepared by bone marrow-derived stromal cells," Proc Natl Acad Sci U S A, 1990 Sep, 87(18):7260-4.	
	C70	VANDERKERKEN K, et al., "Recombinant osteoprotegerin decreases tumor burden and increases survival in a murine model of multiple myeloma," Cancer Res., 2003 Jan 15, 63(2):287-9.	
	C71	WEI S, et al., "Receptor activator of nuclear factor-kappa b ligand activates nuclear factor-kappa b in osteoclast precursors," Endocrinology, 2001 Mar, 142(3):1290-5.	
	C72	WEKERLE T, et al., "Mechanisms of transplant tolerance induction using costimulatory blockade," Curr Opin Immunol., 2002 Oct, 14(5):592-600.	

Examiner Signature		Date Considered	9/5/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Substitute for form 1449A/PTO (Modified)				Complete if Known	
INFORMATION DISCLOSURE STATEMENT BY APPLICANT				Application Number	10/611,363
				Filing Date	July 1, 2003
				First Named Inventor	DESJARLAIS, John R.
				Art Unit	1647
(use as many sheets as necessary)				Examiner Name	~~To Be Assigned~~ R.M. DeBerg
Sheet	6	of	6	Attorney Docket Number	A-71486-2

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[6]
JRD	C73	WILLARD D, et al., "Expression, purification, and characterization of the human receptor activator of NF-kappaB ligand (RANKL) extracellular domain," Protein Expr Purif., 2000 Oct, 20(1):48-57.	
	C74	WONG BR, et al., "TRANCE is a novel ligand of the tumor necrosis factor receptor family that activates c-Jun N-terminal kinase in T cells," J Biol Chem., 1997 Oct 3, 272(40):25190-4.	
	C75	WUYTS W, et al., "Evaluation of the role of RANK and OPG genes in Paget's disease of bone," Bone, 2001 Jan, 28(1):104-7.	
	C76	YAMAGISHI T, et al., "Reciprocal control of expression of mRNAs for osteoclast differentiation factor and OPG in osteogenic stromal cells by genistein: evidence for the involvement of topoisomerase II in osteoclastogenesis," Endocrinology, 2001 Aug, 142(8):3632-7.	
JRD	C77	YASUDA H,et al., "Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL," Proc Natl Acad Sci U S A, 1998 Mar 31, 95(7):3597-602.	

Examiner Signature	R.M. DeBerg	Date Considered	9/5/06

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST. 16 if possible. [6] Applicant is to place a check mark here if English Language Translation is attached.

Notice of References Cited	Application/Control No. 10/611,363	Applicant(s)/Patent Under Reexamination DESJARLAIS ET AL.	
	Examiner Regina M. DeBerry	Art Unit 1647	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	Classification
	A	US-			
	B	US-			
	C	US-			
	D	US-			
	E	US-			
	F	US-			
	G	US-			
	H	US-			
	I	US-			
	J	US-			
	K	US-			
	L	US-			
	M	US-			

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	Wells, Additivity of mutational effects in proteins. Biochemistry 29:8509-8517 (1990).
	V	Ngo et al., The Protein Folding Problem and Tertiary Structure Prediction, pp. 433-440 and 492-495 (1994).
	W	Freshney, Culture of animal cells. A manual of basic technique, pages 3-4. Alan R. Liss, Inc., New York (1983).
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Regina M DeBerry 9/18/06

Biochemistry

Volume 29, Number 37 September 18, 1990

Perspectives in Biochemistry

Additivity of Mutational Effects in Proteins

James A. Wells

Protein Engineering Department, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080.

Received April 19, 1990; Revised Manuscript Received May 29, 1990

The energetics of virtually all binding functions in proteins is the culmination of a set of molecular interactions. For example, removal of a single molecular contact by a point mutation causes relatively small reductions (typically 0.5–5 kcal/mol) in the free energy of transition-state stabilization [for reviews see Fersht (1987) and Wells and Estell (1988)], protein–protein interactions (Laskowski et al., 1983, 1989; Ackers & Smith, 1985), or protein stability [for review see Matthews (1987)] compared to the overall free energy associated with these functional properties (usually 5–20 kcal/mol). Thus, it is possible to modulate protein function by mutation at many contact sites. In fact, to design large changes in function will often require mutation of more than one functional residue.

There is now a large data base for free energy changes that result when single mutants are combined. A review of these data shows that, in the majority of cases, the sum of the free energy changes derived from the single mutations is nearly equal to the free energy change measured in the multiple mutant. However, there are two major exceptions where such simple additivity breaks down. The first is where the mutated residues interact with each other, by direct contact or indirectly through electrostatic interactions or structural perturbations, so that they no longer behave independently. The second is where the mutation causes a change in mechanism or rate-limiting step of the reaction. It is important to note that the additive effects discussed here do not change the molecularity of their respective reactions. When the molecularity of the reaction changes [as in comparing the free energy of binding of one linked substrate (A–B) versus the sum of two fragments (A plus B)], large deviations from simple additivity can result from entropic effects (Jencks, 1981). Although the focus here is on enzyme activity, similar conclusions may be drawn from mutations affecting protein–protein interactions, protein–DNA recognition, or protein stability. Some practical examples and applications are discussed.

Additivity Relationships

The change in free energy of a functional property caused by a mutation at site X is typically expressed relative to that of the wild-type protein as $\Delta\Delta G_{(X)}$. Such free energy changes for two single mutants (X and Y) can be related to those of a double mutant (designated X,Y) by eq 1 (Carter et al., 1984; Ackers & Smith, 1985). The ΔG_I term (also called the coupling energy; Carter et al., 1984) should reflect the extent to which the change in energy of interaction between sites X and Y affects the functional property measured. It is possible for ΔG_I to be either positive or negative depending upon whether the interactions between the mutant side chains reduce or enhance the functional property measured. Furthermore, the ΔG_I term should not exceed the free energy of interaction between side chains at sites X and Y except in cases where these mutations cause large structural perturbations. This was first applied to evaluating the functional independence of residues mutated in tyrosyl-tRNA synthetase (Carter et al., 1984). In one case the sum of the $\Delta\Delta G$ values for single mutants was equal to that of the double mutant, indicating the sites functioned independently; in another example there was a large discrepancy, suggesting the sites were interacting.

$$\Delta\Delta G_{(X,Y)} = \Delta\Delta G_{(X)} + \Delta\Delta G_{(Y)} + \Delta G_I \qquad (1)$$

Simple Additivity in Transition-State Binding Interactions

The strengths of noncovalent interactions are strongly dependent upon the nature of the two groups and the distance (r) between them. For example, the free energy of charge–charge, random charge–dipole, random dipole–dipole, van der Waals attraction, and repulsion decay as $1/r$, $1/r^4$, $1/r^5$, $1/r^6$, and $1/r^{12}$, respectively [for review see Fersht (1985)]. Thus, when the side chains at sites X and Y are remote to one another and assuming no large structural perturbations, the ΔG_I term should be negligible and eq 1 thus simplifies to

$$\Delta\Delta G_{(X,Y)} \simeq \Delta\Delta G_{(X)} + \Delta\Delta G_{(Y)} \qquad (2)$$

This situation, here referred to as simple additivity, is generally observed except where side chains are close to each other or when one or both of the mutants change the rate-limiting step or reaction mechanism. These principles are well illustrated from data of additive mutational effects on transition-state stabilization energies.

0006-2960/90/0429-8509$02.50/0

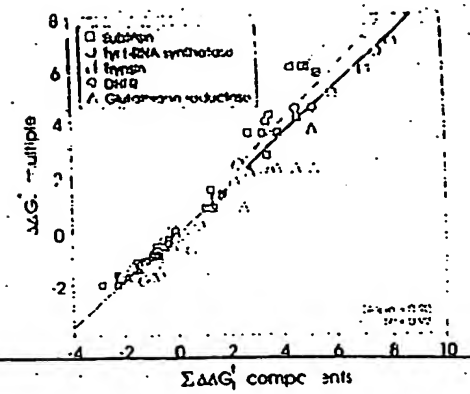

FIGURE 1: Plot of the changes in transition-state stabilization energies for the multiple mutant versus the sum for the component mutants. Data are taken from Table I and represent mutants from subtilisin (□), tyrosyl-tRNA synthetase (○), trypsin (◻), DHFR (○), and glutathione reductase (△), where mutant or wild-type side chains should not contact one another. The dashed line has a slope of 1, and the solid line is a best fit to all the data.

Changes in transition-state stabilization energy ($\Delta\Delta G_T^\ddagger$) caused by a mutation can be calculated from eq 3 (Wilkinson et al., 1983), in which R is the gas constant, T is the absolute

$$\Delta\Delta G_T^\ddagger = -RT \ln \frac{(k_{cat}/K_M)_{mutant}}{(k_{cat}/K_M)_{wild\text{-}type}} \qquad (3)$$

temperature, k_{cat} is the turnover number, and K_M is the Michaelis constant for the mutant and wild-type enzyme against a fixed substrate. $\Delta\Delta G_T^\ddagger$ represents the change in free energy to reach the transition-state complex ($E{\cdot}S^\ddagger$) from the free enzyme and substrate ($E + S$).

To analyze the proposition that the interaction energy term, $\Delta G_{T(I)}^\ddagger$, is relatively small when the sites of mutation (X and Y) are remote to one another, $\Delta\Delta G_T^\ddagger$ values were collected from the literature where side-chain substitutions in the multiple mutant are beyond van der Waals contact (>4 Å distant) from each other (Table I). There are at least 25 examples distributed across five different enzymes where $\Delta\Delta G_T^\ddagger$ values can be calculated for the individual and multiple mutants assayed in at least two different ways. Among these are examples where electrostatic interactions, hydrogen bonding, and steric and hydrophobic effects have been altered separately or in combination with others. The X-ray structures of the wild-type proteins show that the wild-type side chains are not in contact. Modeling suggests the mutant side chains are beyond possible van der Waals contact unless the mutant side chains were to cause significant changes in the overall protein structure. Such large changes are rarely observed in structures of site-specific mutant proteins (Katz & Kossiakoff, 1986; Alber et al., 1987; Howell et al., 1986; Wilde et al., 1988) or even highly variant natural proteins (Chothia & Lesk, 1986).

A collective plot of the sum of the $\Delta\Delta G_T^\ddagger$ values for the component mutants versus the corresponding multiple mutant (Table I) gives a remarkably strong correlation ($R^2 = 0.92$) with a slope near unity (Figure 1). The simplest interpretation is that the interaction term, $\Delta G_{T(I)}^\ddagger$, is small compared to the overall effects on $\Delta\Delta G_{T(X,Y)}^\ddagger$. It is formally possible that there are large and compensating effects between side chains X and Y that systematically lead to small net values for $\Delta G_{T(I)}^\ddagger$.

There are some notable exceptions that weaken the correlation within the data set (Table I). In particular, combining the R204L mutation in *Escherichia coli* glutathione reductase gives a less than additive effect, especially when combined with another mutant, R198M (Scrutton et al., [illegible] residues are not in direct contact, but [illegible] a salt bridge with the 2′-phosphate group [illegible] the largest discrepancies are when these mutants [illegible] with NADPH as compared to NADH. Similarly, the sum of the $\Delta\Delta G_T^\ddagger$ values for two positively charged component mutants in subtilisin (D99K and E156K) overestimates the effect of the multiple mutant when assayed with an Arg but not with a Phe substrate (Russell & Fersht, 1987). Such discrepancies are not too surprising because charge–charge interactions fall off as $1/r$ and can exhibit long-range effects in proteins [for example, see Russell and Fersht (1988)]. The physical basis for other large discrepancies not involving electrostatic substitutions is less clear but may involve unexpectedly large structural changes or changes in enzyme mechanism (see below).

These additivity tests are not particularly dominated by one of the single mutants in the sum. The average contribution (±SE) for the most dominant mutant in each sum calculated from the 69 additivity tests given in Table I is only 68% (±15%) of the total sum (theoretical is ~50%). Furthermore, the plot in Figure 1 is not analogous to graphs of correlated variables, where A is plotted versus the sum of A + B, because in Figure 1 the values on the y-axis are determined *independently* from those on the x-axis.

Complex Additivity in Transition-State Stabilization—When $\Delta G_{T(I)}^\ddagger \neq 0$

(A) Change in Interaction Energy between Sites X and Y. Where residues X and Y are close enough to contact, it is more likely that the $\Delta G_{T(I)}^\ddagger$ term will be significant. There are 11 examples collectively from tyrosyl-tRNA synthetase and subtilisin that fit this category (Table II).

A series of mutants in tyrosyl-tRNA synthetase at positions 48 and 51 (Carter et al., 1984; Lowe et al., 1985) show complex additivity (Table II). His48 and Thr51 in the wild-type structure are next to each other on adjacent turns of an α-helix. His48 hydrogen bonds to the ribose ring oxygen of ATP while Thr51 can make van der Waals contact with ATP. The T51P mutation increases the catalytic efficiency of the enzyme in some assays by more than −2 kcal/mol (Wilkinson et al., 1984). However, when this mutation is combined with mutations at position 48, the effects are not simply additive. An X-ray structure of the T51P mutant indicates there are no structural changes in the α-helix (Brown et al., 1987). Instead, it is suggested that the T51P mutant is improved over wild type because the wild-type enzyme contains a bound water in the vicinity of Thr51 that disfavors substrate binding. Blow and co-workers (Brown et al., 1987) argue that the change in solvent structure propagated to position 48 may account for the complex additivity. In the previous section, the double mutant (H48G,T51A) exhibited nearly simple additivity (Table I). Presumably, the smaller and less hydrophobic alanine substitution at position 51 should not introduce as large a change in solvent structure as the pyrrolidone ring of proline.

In the case of subtilisin (Table II), Glu156 is near the top of the P1 binding crevice while Gly166 is at the bottom. In the wild-type enzyme these sites do not make direct van der Waals contact, but large side chains substituted at position 166 can be modeled to contact the residue at position 156. In fact, X-ray structural analysis shows that an Asn side chain at position 166 makes a good hydrogen bond with Glu156 (Bott et al., 1987). Moreover, all of the substitutions are polar or charged, the energetics of which are expected to be the most long range. Thus, the mutant side chains alter substantially the intramolecular interactions between positions 156 and 166.

Table I: Comparison of Sums of $\Delta\Delta G_T$[a] from Component Mutants vs the Multiple Mutant Where the Mutant or Wild-Type Side Chains *Do Not* Contact One Another

assay	$\Delta\Delta G_T$[a] component mutants			sum	multiple mutant
Tyrosyl-tRNA Synthetase					
	C35G + H48G[a]				
ATP/PP_i	+1.20	+1.04		+2.24	+2.30
ATP/tRNA	+1.05	+1.13		+2.18	+1.68
Tyr/PP_i	+1.14	+1.12		+2.26	+2.32
Tyr/tRNA	+0.32	+1.12		+1.45	+1.20
	C35G + T51P				
ATP/PP_i	+1.20	−1.91		−0.71	−1.14
ATP/tRNA	+1.05	−2.35		−1.30	−1.68
Tyr/PP_i	+1.14	−0.64		+0.50	−0.74
Tyr/tRNA	+0.32	+0.50		+0.82	+0.21
	C35G + T51C[a]				
ATP/tRNA	+1.05	−0.93		+0.12	−0.22
ATP/Tyr	+1.14	−0.91		+0.23	−0.13
	H48N + T51A[c]				
ATP/PP_i	+0.26	−0.46		−0.12	+0.04
ATP/tRNA	−0.13	−0.32		−0.45	−0.37
	T40A + H45G[d]				
Tyr/Tyr	+5.02	+3.15		+8.17	+6.95
ATP/Tyr	+5.13	+2.44		+7.57	+6.67
Rat Trypsin					
	G216A + G226A[e]				
K	+2.75	+3.13		+5.88	+5.07
R	+2.19	+4.91		+7.10	+5.90
Dihydrofolate Reductase ($\Delta\Delta G_{binding}$)					
	F31V + L54G[f]				
H_2F	+1.6	+2.9		+4.5	+4.5
MTX	+2.2	+2.9		+5.1	+4.5
Subtilisin BPN′					
	E156S + Y217L + G169A[g]				
E	−1.43	−0.87	−0.62	−2.92	−2.06
Q	−0.60	−0.36	−0.32	−1.28	−1.14
A	−0.15	−0.41	−0.27	−0.83	−0.92
K	+1.70	−0.08	−0.30	+1.32	+0.87
M	−0.86	−0.32	−0.39	−1.57	−1.41
F	−0.61	−0.29	−0.66	−1.56	−1.17
Y	−0.24	−0.12	−0.41	−0.77	−0.59
	E156S + Y217L				
E	−1.43	−0.87		−2.30	−1.67
Q	−0.60	−0.36		−0.96	−0.96
A	−0.15	−0.41		−0.56	−0.53
K	+1.70	−0.08		+1.62	+1.33
M	−0.86	−0.32		−1.18	−1.11
F	−0.61	−0.29		−0.90	−0.84
Y	−0.24	−0.12		−0.36	−0.32
	E156S, Y217L + G169A				
E	−1.67	−0.62		−2.29	−2.06
Q	−0.96	−0.32		−1.28	−1.14
A	−0.53	−0.27		−0.80	−0.92
K	+1.33	−0.30		+1.03	+0.87
M	−1.11	−0.39		−1.50	−1.41
F	−0.84	−0.66		−1.50	−1.17
Y	−0.32	−0.41		−0.73	−0.59
	D99S + E156S[h]				
R	+0.47	+0.77		+1.24	+1.52
F	0	−0.62		−0.62	−0.52
Subtilisin BPN′					
	D99K + E156K				
R	+1.29	+2.12		+3.41	+2.74
F	+0.13	−0.49		−0.36	−0.42
	G166A + E156S, G169A, Y217L[i]				
F	−0.40	−1.46		−1.86	−1.76
Y	+0.94	−1.03		−0.09	+0.02
	G166A + S24C, H64A				
F	−0.40	+4.96		+4.56	+4.11
Y	+0.94	+4.40		+5.34	+5.84
	E156S, G169A, Y217L + S24C, H64A				
F	−1.46	+4.96		+3.50	+4.21
Y	−1.03	+4.40		+3.37	+3.96
	S24C, H64A, G169A, Y217L + G166A				
F	+4.21	−0.40		+3.81	+3.53
Y	+3.96	+0.94		+4.90	+6.07
	S24C, H64A, G166A + E156S, G169A, Y217L				
F	+4.11	−1.46		+2.65	+3.53
Y	+5.84	−1.03		+4.81	+6.07
	S24C, H64A + E156S, G166A, G169A, Y217L				
F	+4.96	−1.76		+3.20	+3.53
Y	+4.40	+0.02		+4.42	+6.07
***E. coli* Glutathione Reductase**					
	A179G + R198M[j]				
NADH	−1.10	−0.62		−1.72	−1.32
NADPH	+0.08	+2.68		+2.76	+2.11
	A179G + R204L				
NADH	−1.10	+0.41		−0.69	−1.54
NADPH	+0.08	+2.42		+2.50	+0.87
	R198M + R204L				
NADH	−0.62	+0.41		−0.21	−0.51
NADPH	+2.68	+2.42		+5.10	+3.70
	A179G + R179M, R204L				
NADH	−1.10	−0.51		−1.61	−1.72
NADPH	+0.08	+3.70		+3.78	+2.22
	R198M + A179G, R204L				
NADH	−0.62	−1.54		−2.16	−1.72
NADPH	+2.68	+0.87		+3.55	+2.22
	R204L + A179G, R198M				
NADH	+0.41	−1.32		−0.91	−1.72
NADPH	+2.42	+2.11		+4.53	+2.22
	R179G + R198M + R204L				
NADH	−1.10	−0.62	+0.41	−1.31	−1.72
NADPH	+0.08	+2.68	+2.42	+5.18	+2.22

[a] Carter et al. (1984). The assays refer to measurements of ATP-dependent pyrophosphate exchange (ATP/PP_i) or tRNA charging (ATP/tRNA) under saturating conditions for tyrosine and vice versa for Tyr/PP_i exchange and Tyr/tRNA charging. [b] Lowe et al. (1985). The ATP/Tyr activation assay refers to formation of tyrosyl adenylate under saturating concentrations of tyrosine. [c] Jones et al. (1986). [d] Leatherbarrow et al. (1986). The ATP/Tyr and Tyr/Tyr activation assays refer to formation of tyrosyl adenylate under pre-steady-state conditions, and k_{cat}/K_M is calculated from k_3/K_S for tyrosine and ATP, respectively. [e] Craik et al. (1985). The substrate was D-Val-Leu-(X)-aminofluorocoumarin where the P1 residue (X) is either Lys (K) or Arg (R). [f] Mayer et al. (1986). The ligand was either dihydrofolate (H_2F) or methotrexate (MTX). [g] Wells et al. (1987a). The substrate was succinyl-L-Ala-L-Ala-L-Pro-L-(X)-*p*-nitroanilide where the P1 (X) residue (Schechter & Berger, 1967) was either Glu (E), Gln (Q), Ala (A), Lys (K), Met (M), Phe (F), or Tyr (Y). [h] Russell and Fersht (1987). The substrate was benzoyl-L-Val-Gly-L-Arg-*p*-nitroanilide (R) or succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (F). [i] Carter et al. (1989). The substrate was succinyl-L-Phe-L-Ala-L-His-L-(X)-*p*-nitroanilide where X was either Phe (F) or Tyr (Y). [j] Scrutton et al. (1990). The assay followed the reduction of oxidized glutathione by NADH or NADPH.

Table II: Comparison of Sums of $\Delta\Delta G_T^\ddagger$ from Component Mutants vs the Multiple Mutant Where the Mutant Side Chains *Can* Contact One Another

assay[a]	$\Delta\Delta G_T^\ddagger$ component mutants	$\Delta\Delta G_T^\ddagger$ sum	$\Delta\Delta G_T^\ddagger$ multiple mutant
Tyrosyl-tRNA Synthetase			
H48G + T51P[b]			
ATP/PP$_i$	+1.04 −1.91	−0.87	+1.07
ATP/tRNA	+1.13 −2.35	−1.22	+0.77
Tyr/PP$_i$	+1.12 −0.64	+0.48	+1.02
Tyr/tRNA	+1.12 +0.50	+1.63	+0.17
ATP/Tyr	+0.95 −1.99	−1.04	+1.04
Tyr/ATP	+1.07 −0.38	+0.69	+0.82
H48N + T51P			
ATP/Tyr	+0.18 −1.99	−1.81	−0.76
Tyr/Tyr	+0.36 −0.38	−0.02	−0.64
ATP/tRNA	−0.02 −2.23	−2.25	−1.07
N48G + T51P			
ATP/Tyr	+0.37 −0.94	−0.57	+0.86
Tyr/Tyr	+0.41 −1.00	−0.59	+0.45
ATP/tRNA	+1.26 −1.05	+0.21	+0.90
Q48G + T51P			
ATP/Tyr	−1.31 −1.09	−2.40	−1.22
Tyr/Tyr	−2.05 −1.65	−3.70	−2.31
ATP/tRNA	−1.87 −1.85	−3.72	−2.23
H48Q + T51P			
ATP/Tyr	+2.26 −1.99	+0.27	+1.17
Tyr/Tyr	+3.13 −0.38	+2.75	+1.48
ATP/tRNA	+3.11 −2.23	+0.88	+1.25
Subtilisin BPN'			
E156Q + G166D[d]			
Q	−1.04 +1.27	+0.23	+0.75
M	−0.45 +1.83	+1.38	+0.16
K	+2.15 +0.53	+2.68	+0.26
E156S + G166D			
Q	−0.59 +1.27	+0.68	+0.74
M	−0.85 +1.83	+0.98	+0.66
K	+1.68 +0.53	+2.22	+0.49
E156Q + G166N			
E	−1.71 −0.11	−1.82	−0.69
Q	−1.04 +0.14	−0.90	−0.77
M	−0.45 +0.18	−0.27	−1.10
K	+2.15 +0.48	+2.73	+1.15
E156S + G166N			
E	−1.44 −0.11	−1.55	−0.51
Q	−0.59 +0.14	−0.45	−0.85
M	−0.85 +0.18	−0.67	−0.78
K	+1.68 +0.48	+2.16	+1.26
E156S + G166K			
E	−1.44 −3.49	−4.93	−4.49
Q	−0.59 −1.03	−1.62	−0.95
M	−0.85 −1.37	−2.22	−1.12
K	+1.68 +0.51	+2.19	+1.88
E156Q + G166K			
E	−1.71 −3.49	−5.20	−4.49
Q	−1.04 −1.03	−2.07	−0.95
M	−0.45 −1.37	−1.82	−1.12
K	+2.15 +0.51	+2.66	+1.88

[a] See Table I for description assays. [b] Lowe et al. (1985). [c] Carter et al. (1984). [d] Wells et al. (1987b).

In these six examples there are large and systematic discrepancies between the sum of the $\Delta\Delta G_T^\ddagger$ values for the single mutants and those of the corresponding double mutant (Wells et al., 1987b). In almost all cases, the sum of the $\Delta\Delta G_T^\ddagger$ values for the single mutants is much greater than the value for the multiple mutant. Nonetheless, the $\Delta\Delta G_T^\ddagger$ value predicted from the sum of the single mutants does have the same sign as that for the double mutant, so that the single mutants predict qualitatively the effect on the multiple mutant.

A plot (Figure 2) of the collective data set from Table II is in contrast to that seen in Figure 1. The $\Delta\Delta G_T^\ddagger$ values for the multiple mutants correlate more poorly with the sum of

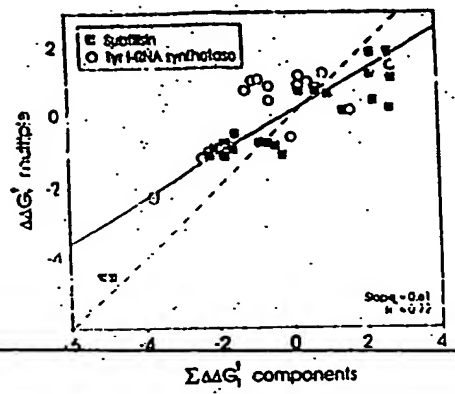

FIGURE 2: Data are taken from Table II for mutants of subtilisin (■) or tyrosyl-tRNA synthetase (○) where mutant or wild-type side chains can contact each other. The dashed line represents a theoretical line of unity slope, and the solid line represents the best fit.

the component single mutants ($R^2 = 0.72$). Moreover, the slope of the line (0.61) is much below unity. This indicates that the function of one residue is compromised by mutation of another. Of the 40 additivity examples, the average contribution of the most dominant single mutant to the sum of the $\Delta\Delta G_T^\ddagger$ values is 71% (±13%) of the total. Thus (as in Figure 1), both single mutants can contribute substantially to free energy changes measured in the multiple mutant. However, this data set is derived from mutations at only two different sites on two different proteins.

In summary, complex additivity can be observed when mutations at sites X and Y change the intramolecular interaction energy between sites. This can be mediated by direct steric, electrostatic, hydrogen-bonding, or hydrophobic interactions or indirectly through large structural changes in the protein, solvent shell, or electrostatic interactions. Complex additivity is most likely to occur where the sites of mutation are very close together and larger or chemically divergent side chains are introduced.

(*B*) *Mutations at Sites X or Y Change the Enzyme Mechanism or Rate-Limiting Step.* If the catalytic functions of two or more residues are interdependent, then a mutation of one residue can affect the functioning of the other(s). This form of complex additivity is well illustrated for mutations in the catalytic triad and oxyanion binding site of subtilisin (Carter & Wells, 1988, 1990). In the catalytic mechanism of subtilisin (Figure 3), the rate-limiting step in amide bond hydrolysis is transfer of the proton from Ser221 to His64 with nucleophilic attack upon the scissile carbonyl carbon. This is accompanied by electrostatic stabilization of the protonated imidazole by Asp32 and hydrogen bonding to the oxyanion by the side chain of Asn155 and the main-chain amide of Ser221. Mutational analysis shows that once the catalytic Ser221 is mutated to Ala (S221A), additional mutations in the triad or oxyanion binding site cause no further loss in catalytic efficiency (Table III).

The S221A enzyme retains a catalytic activity that is still 10^4 above the solution hydrolysis rate (Carter & Wells, 1988). It is proposed that this residual activity is derived from remaining transition-state binding contacts outside of the catalytic triad coupled with solvent attack upon the carbonyl carbon from the face opposite position 221 (Carter & Wells, 1990). This proposal is based on a model showing that there is no room for a water molecule near Ala221 once the substrate is bound. Furthermore, conversion of Asn155 to Gly enhances the activity of the S221A mutant by −1.2 kcal/mol (Table III).

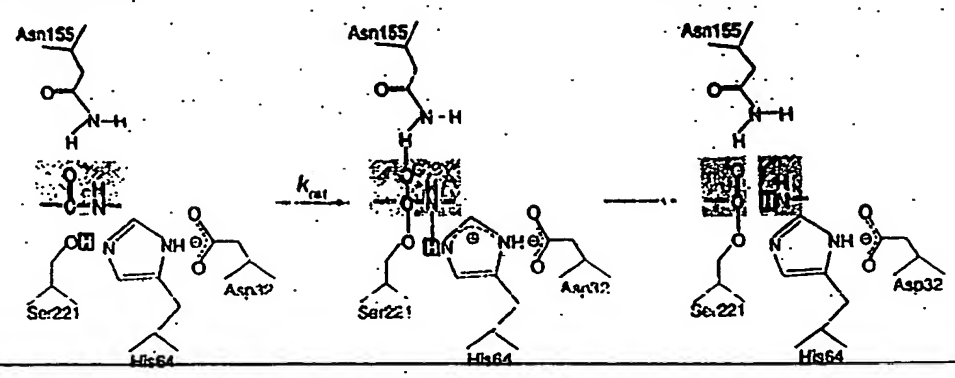

FIGURE 3: Schematic diagram of the mechanism of subtilisin showing the rate-limiting activation step for hydrolysis of peptide bonds. Reproduced with permission from Carter and Wells (1988). Copyright 1988 Macmillan.

Table III: Comparison of Sums of $\Delta\Delta G_T^{\ddagger}$ from Component Mutants vs the $\Delta\Delta G_T^{\ddagger}$ for Multiple Mutants in the Catalytic Triad and Oxyanion Binding Site of Subtilisin BPN'[a]

component mutants			sum	multiple mutant
S221A + H64A[b]				
+8.93	+8.84		+17.76	+8.83
S221A + D32A				
+8.93	+6.52		+15.45	+8.86
H64A + D32A				
+8.84	+6.52		+15.36	+7.48
S221A + H64A + D32A				
+8.93	+8.84	+6.52	+24.29	+8.65
S221A + H64A, D32A				
+8.93	+7.48		+16.40	+8.65
H64A + S221A, D32A				
+8.84	+8.86		+17.70	+8.65
D32A + S221A, H64A				
+6.52	+8.83		+15.35	+8.65
S221A + N155G[c]				
+8.93	+3.08		+12.01	+7.70

[a] All enzymes were assayed with the substrate succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide. [b] Carter and Wells (1988). [c] Carter and Wells (1990).

This is consistent with the opposite-face solvent attack mechanism of S221A, because the oxyanion (Figure 3) would develop away from Asn155 and the N155G mutation improves solvent accessibility to the scissile carbonyl carbon.

Complex additivity is also seen for subtilisin mutated at positions 64 and 32. The double (H64A,D32A) and corresponding single mutants show a linear dependence upon hydroxide ion concentration (between pH 8 and 10) that may reflect hydroxide assistance in the deprotonation of the Oγ of Ser221 (Carter & Wells, 1988). Thus, once His64 is converted to Ala, Asp32 is a liability, presumably by electrostatic repulsion of hydroxide ion. [Note the −1.3 kcal/mol improvement in $\Delta\Delta G_T^{\ddagger}$ for the double mutant (H64A,D32A) compared to H64A alone; Table III.]

In summary, if an enzyme mechanism relies upon cooperative interaction between two or more residues, then multiple mutations within this subset can result in large values for $\Delta G_{T(i)}^{\ddagger}$. In fact, if the mechanism is changed substantially, residues that were a catalytic asset can become a liability. Simple additivity can also break down when one or more of the mutations cause a change in the rate-limiting step. In an extreme case, one may have a number of mutants in an enzyme that enhance the activity, but the cumulative enhancement of activity could not go beyond the diffusion-controlled limit (Albery & Knowles, 1976).

Additive Effects on Substrate Binding

The analysis above considered changes in binding free energies between the free enzyme and substrate (E + S) to yield the bound transition-state complex (E·S‡). The steady-state kinetic analysis for subtilisin and tyrosyl-tRNA synthetase is such that the K_M values approximate the enzyme–substrate dissociation constant K_s. Additivity analysis based on calculations of $\Delta\Delta G_{binding}$ (from K_M values) or $\Delta\Delta G_{cat}$ (from k_{cat} values) yields qualitatively the same results (not shown) as shown in Tables I and II and Figures 1 and 2. Thus, deviations from simple additivity are not systematically found in either the energetics to form the E·S complex or those to reach E·S‡.

Additive Effects on Protein–Protein Interactions

The first clear examples of additive binding effects caused by amino acid replacements in proteins were reported by Laskowski et al. (1983) and reviewed by others (Ackers & Smith, 1985; Horovitz & Rigbi, 1985). One hundred natural variants of a proteinase inhibitor, the ovomucoid third domain, have been isolated and sequenced from the eggs of different bird species (Empie & Laskowski, 1982; Laskowski et al., 1987). This is a nested set of proteins because for any one of these avian inhibitors there is a close relative containing only one or a few amino acid substitutions. Moreover, the association constants (K_a) of these inhibitors with a variety of serine proteinases vary over an enormous range (10^9-fold). Laskowski et al. (1983, 1989) have shown that the effect of a given residue replacement on K_a is about the same irrespective of the inhibitor scaffold the replacement is made in.

In addition to ovomucoid, four additivity examples have been constructed from natural variants at the subunit interface of tetrameric hemoglobin (Ackers & Smith, 1985). Three additivity examples have been analyzed for interactions of hGH with its receptor (B. C. Cunningham and J. A. Wells, unpublished results) and one example for association of synthetic variants of the RNase S peptide with RNase S protein (Mitchinson & Baldwin, 1986). The entirety of this data set is not tabulated because much on the ovomucoid inhibitors and hGH is unpublished. Nonetheless, these researchers were kind enough to provide their data formatted so it could be plotted collectively in Figure 4. These data consist of 91 additivity examples (80 in ovomucoids alone), representing 22 multiple mutants across four different proteins, and span a wide range of change in binding free energy (−10 to +7

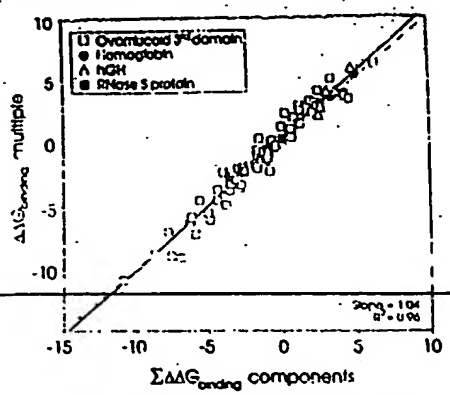

FIGURE 4: Plot showing the sum of changes in free energies of binding at protein–protein interfaces for component mutants versus the corresponding multiple mutant. Data represent interactions between ovomucoid third domain and various serine proteases (□) (R. Wynn and M. Laskowski, personal communication), regulatory interface of $\alpha_2\beta_2$ hemoglobin (●) (Ackers & Smith, 1985), hGH and its receptor (stippled Δ) (B. Cunningham and J. Wells, personal communication), and RNase S peptide and S protein (▣) (Mitchinson & Baldwin, 1986). The dashed line represents a line of unity slope, and the solid line is the best fit.

kcal/mol). The plot shows a very strong linear correlation (R^2 = 0.96) with a slope near unity. Although the data for the ovomucoid were not sorted to evaluate changes at intramolecular contact sites, most are not expected to be in contact, and all of the other examples represent noncontact sites. Thus, the large data base derived from natural variants of ovomucoid third domain, as well as a smaller number of examples from several other proteins, indicates that multiple mutations at protein–protein interfaces commonly produce simple additive effects.

Additive Effects in DNA–Protein Interactions

One of the clear advantages in analyzing DNA–protein interactions is the ability to apply powerful selections that make analysis by random mutational studies feasible. Additivity in DNA–protein interactions was first demonstrated by reversion analysis of λ repressor (Nelson & Sauer, 1985). A mutation that decreased the binding affinity for the λ operator site (K4Q) was reverted by mutations at several second sites (E34K, G48S, and E83K). When these second-site revertants were introduced into wild-type λ repressor, they caused increases in affinity similar to those observed in the first-site suppressor mutant (K4Q).

Functional independence for mutations at DNA–protein contacts has been demonstrated by additive effects for mutants of CAP (catabolite gene activator protein) and its operator sequence (Ebright et al., 1987) as well as *lac* repressor and its corresponding operator sequence (Ebright, 1986). Simple additivity of mutational effects in the operator sequences for Cro repressor (Takeda et al., 1989) and λ repressor (Sarai & Takeda, 1989) has been most systematically demonstrated. Simple additivity has also been reported for multiple mutations in the *lac* repressor (Lehming et al., 1990). In fact, simple additivity is so predictable in DNA–protein interactions that the observation of complex additivity has been used to predict specific DNA–protein contacts in the *lac* repressor–operator complex (Ebright, 1986).

Additive Effects on Protein Stability

The first systematic analysis of additive effects of site-specific mutations on protein stability was reported by Shortle and Meeker (1986). Five multiple mutants in staphylococcal nuclease were constructed from a group of random single mutants that were screened initially for their ability to affect the stability of the enzyme in vivo. The component mutants do not make direct contact with each other in the multiple mutants. Generally, these variants exhibit nearly additive effects except for the double mutant V66L,G88V (Table IV). In addition to those of staphylococcal nuclease, additive effects on the $\Delta\Delta G_{unfolding}$ (assayed by reversible denaturation) have also been determined for the N-terminal domain of λ repressor (one example; Hecht et al., 1986), the α-subunit of *E. coli* Trp synthetase (one example; Hurle et al., 1986), T4 lysozyme (six examples; Wetzel et al., 1988), the gene V product of bacteriophage f1 (one example; Sandberg & Terwilliger, 1989),

Table IV: Comparison of Sums of $\Delta\Delta G_{unfolding}$ from Component Mutants vs the Multiple Mutant

assay	$\Delta\Delta G_{unfolding}$ component mutants			sum	multiple mutant
Staphylococcal Nuclease					
	V66L + G79S[a]				
GuHCl	−0.2	−2.6		−2.8	−3.3
urea	+0.2	−2.9		−2.7	−3.6
	V66L + G88V				
GuHCl	−0.2	−1.0		−1.2	−2.1
urea	+0.2	−0.9		−0.7	−1.4
	I18M + A90T				
GuHCl	−0.5	−2.7		−3.3	−2.8
urea	−0.7	−2.9		−3.6	−3.8
	I18M + A90S				
GuHCl	−0.6	−1.4		−2.0	−2.2
urea	−0.7	−1.4		−2.1	−2.2
	V66L + G79S + G88V				
GuHCl	−0.2	−2.6	−1.0	−3.8	−3.0
urea	+0.2	−2.9	−0.9	−3.6	−3.4
N-Terminal Domain of λ Repressor					
	G46A + G48A[b]				
thermal melt	+0.7	+0.9		+1.6	+1.1
T4 Lysozyme					
	I3C + C54V[c]				
thermal melt	+1.2	−0.7		+0.5	+0.4
	I3C + C54T				
thermal melt	+1.2	+0.3		+1.5	+1.5
	I3C + C54T + R96H				
thermal melt	+1.2	+0.3	−2.8	−1.3	−2.5
	I3C,C54T + R96H				
thermal melt	+1.5	−2.8		−1.3	−2.5
	I3C + C54T + A146T				
thermal melt	+1.2	+0.3	−1.5	0	−0.5
	I3C,C54T + A146T				
thermal melt	+1.5	−1.5		0	−0.5
Bacteriophage f1 Gene V					
	V35I + I47V[d]				
GuHCl	−0.4	−2.4		−2.8	−2.9
Kringle-2 of tPA					
	H64Y + R68G[e]				
thermal melt	+2.9	+0.7		+3.6	+3.4
Turkey Ovomucoid Third Domain					
	G32A + N28S[f]				
thermal melt	+0.8	−0.5		+0.3	+0.2
	Y20H + N45-CHO				
thermal melt	−0.8	+0.3		−0.5	−0.6
α Subunit of *E. coli* Trp Synthetase					
	Y175C + G211E[g]				
GuHCl	−0.1	+0.3		+0.2	−1.3

[a] Shortle and Meeker (1986). [b] Hecht et al. (1986). [c] Wetzel et al. (1988). [d] Sandberg and Terwilliger (1989). [e] R. Kelley, personal communication. [f] Otlewski and Laskowski (1990). N45-CHO refers to a glycosylation of Asn45. [g] Hurle et al. (1986).

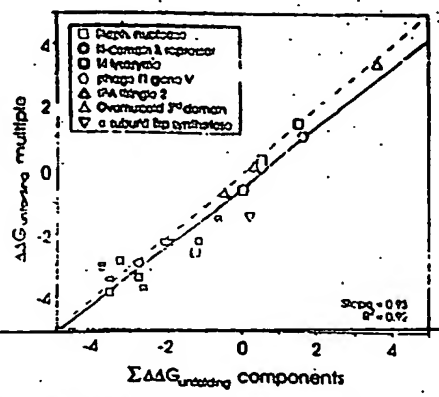

FIGURE 5: Plot showing sum of changes in free energy of unfolding of component mutants and resulting multiple mutant. Data are taken from Table IV and represent staphylococcal nuclease (□), N-terminal domain of λ repressor (○), T4 lysozyme (◻), bacteriophage f1 gene V product (○), Kringle-2 domain of tissue plasminogen activator (△), turkey ovomucoid third domain (△), and the α-subunit of Trp synthetase (▽). The dashed line represents a theoretical line of unity slope, and the solid line represents the best fit.

natural variants of ovomucoid third domain (two examples; Otlewski & Laskowski, 1990), and the Kringle-2 domain of human tissue plasminogen activator (t-PA) (one example; R. Kelley, personal communication).

Collectively, this data set gives a high linear correlation (R^2 = 0.94) and slope near unity (Figure 5). The generally simple additive behavior is somewhat surprising given the highly cooperative nature of protein folding. There are discrepancies in some of the additivity examples besides the staphylococcal nuclease mutant (V66L,G88V). For example, the 1.5 kcal/mol discrepancy for the Y175C,G271E double mutant in Trp synthetase (Table IV) is proposed to result from the fact that these residues are in direct contact (Hurle et al., 1986). Furthermore, proximity effects may account for the large differences between the sum of the component mutants and the multiple mutants for the α-helical double glycine mutant G46A,G48A in λ repressor (Hecht et al., 1986), and when combining R96H with the C3–C97 disulfide mutant in T4 lysozyme (Wetzel et al., 1988). In contrast, an exchange of two side chains that contact one another (V35I and I47V) in the hydrophobic core of the gene V product of f1 phage produced simple additive effects (Sandberg & Terwilliger, 1989; Table IV). It should be noted that this data base exhibiting simple additivity may be biased for single mutants that stably fold, because severely unstable proteins are more difficult to express.

By analogy to transition-state binding effects, one can certainly imagine instances where the stabilizing effects of mutations should reach a plateau. For example, denaturation at high temperatures can become controlled by a chemical step such as deamidation (Ahern et al., 1987), so that additional mutants that stabilize the folded form of the protein may be irrelevant. Another obvious example where complex additivity can be observed in protein stability is the stabilizing effect of disulfide bonds and noncovalent intramolecular contacts that require interactions between two or more residues. In these cases, the stabilizing interaction between two side chains can be broken with only one mutation.

Applications of Additivity in Rational Protein Design

A strategy of additive mutagenesis, where a series of single mutants each making a small improvement in function are combined, is one of the most powerful tools in designing functional properties in proteins. This approach has been remarkably successful in stabilizing proteins to irreversible inactivation, such as λ repressor (Hecht et al., 1986), subtilisin (Bryan et al., 1987; Cunningham & Wells, 1987; Pantoliano et al., 1989), kanamycin nucleotidyltransferase (Liao et al., 1986; Matsumura, 1986), neutral protease (Imanaka et al., 1986), and T4 lysozyme (Wetzel et al., 1988; Matsumura et al., 1989). This strategy has been applied to enhancing the catalytic efficiency of a weakly active variant of subtilisin (Carter et al., 1989), engineering the substrate specificity of subtilisin (Wells et al., 1987a,b; Russell & Fersht, 1987) and the coenzyme specificity of glutathione reductase (Scrutton et al., 1990), designing protease inhibitors with exquisite protease specificity (Laskowski et al., 1989), and recruiting human prolactin to bind to the hGH receptor (Cunningham et al., 1989). In addition, additivity principles have been used to engineer the pH profile of subtilisin (Russell & Fersht, 1987) and to design the affinity and specificity of λ repressor (Nelson & Sauer, 1985).

For this approach to work does not require that all the component mutants act in a simply additive manner but just that their effects accomulate. For example, despite the complex additivity of effects in the catalytic triad of subtilisin, there are mutagenic pathways that are energetically cumulative for installing the triad (Carter & Wells, 1988; Wells et al., 1987c). Starting with the triple mutant S221A,H64A,D32A, there is a progressive enhancement for installing Ser221 (−1.1 kcal/mol), then His64 (−1.0 kcal/mol), and finally Asp32 (−6.5 kcal/mol). Another cumulative pathway of Ser221, then Asp32, and finally His64 is possible if the Ser221,Asp32 intermediate were to use HisP2 substrates (Carter & Wells, 1987). Elaborating such cumulative pathways is important for understanding how a catalytic apparatus may have evolved and is practically useful for considering how to install such catalytic machinery into weakly active catalytic antibodies.

Conclusions

In the majority of cases, combination of mutations that affect substrate or transition-state binding, protein–protein interactions, DNA–protein recognition, or protein stability exhibits simple additivity. Simple additivity is commonly observed for distant mutations at rigid molecular interfaces such as in protein–protein and DNA–protein interactions, where the mutations are unlikely to alter grossly the structure or mode of binding.

Large deviations from simple additivity can occur when the sites of mutations strongly interact with one another (by making direct contact or indirectly through electrostatic interactions or large structural perturbations) and/or when both sites function cooperatively (as for the catalytic triad and oxyanion binding site of subtilisin). Changes at sites that can contact each other do not always lead to complex additivity; this may reflect relatively weak interactions between the two sites or indicate that the interactions are compensatory and appear to be weak.

It is important to point out the magnitude of errors in predicting the free energy effect in the multiple mutant from the component single mutants. Generally, for those cases exhibiting simple additivity (Figures 1, 4, and 5), the discrepancy in free energy between the sums of the components and multiple mutants is about ±25%. Part of this is the result of compounding errors when summing the single mutants, and the rest is presumably due to weak interaction terms. Nonetheless, this means that if the total free energy change is about 3 kcal/mol, the change in the equilibrium constant

(related by $K_{xy}/K_{wt} = 10^{-\Delta\Delta G/RT} = 155$) will often be off by a factor of 4. Thus, while the free energy effects accumulate, significant deviations will occur in predicting the final equilibrium constants when component mutants contribute a large free energy term.

Simple additivity reflects the modularity of component amino acids in protein function. This results from the fact that the perturbations in energetics and structure resulting from most mutations are highly localized. In the past six years, an additive mutagenesis strategy has been extremely effective in engineering proteins. Of course, nature has been using this strategy much longer.

Acknowledgments

I am grateful to Dr. Michael Laskowski and Rich Wynn for providing their data prior to publication on ovomucoid third domain and similarly to Brian Cunningham, Paul Carter, and Robert Kelley for making available their unpublished data. I am indebted for useful discussions with Drs. Michael Laskowski, Paul Carter, Jack Kirsch, and Tony Kossiakoff and many of my colleagues at Genentech and to Drs. Richard Ebright and William Jencks and those above for critical reading of the manuscript.

Registry No. RNase, 9001-99-4; tyrosyl-tRNA synthetase, 9023-45-4; trypsin, 9002-07-7; dihydrofolate reductase, 9002-03-3; subtilisin BPN', 9014-01-1; glutathione reductase, 9001-48-3; staphylococcal nuclease, 9013-53-0; lysozyme, 9001-63-2; plasminogen activator, 105913-11-9; tryptophan synthetase, 9014-52-2.

References

Ackers, G. K., & Smith, F. R. (1985) *Annu. Rev. Biochem. 54*, 597–629.

Ahern, T. J., Casal, J. I., Petsko, G. A., & Klibanov, A. M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 675–679.

Alber, T., Dao-pin, S., Wilson, K., Wozniak, J. A., Cook, S. P., & Matthews, B. W. (1987) *Nature 330*, 41–46.

Albery, W. J., & Knowles, J. R. (1976) *Biochemistry 15*, 5631–5640.

Ardelt, W., & Laskowski, M., Jr. (1990) *J. Mol. Biol.* (submitted for publication).

Bott, R., Ultsch, M., Wells, J., Powers, D., Burdick, D., Struble, M., Burnier, J., Estell, D., Miller, J., Graycar, T., Adams, R., & Power, S. (1987) *ACS Symposium Series 334* (LeBaron, H. M., Mumma, R. O., Honeycutt, R. C., & Deusing, J. H., Eds.) pp 139–147, American Chemical Society, Washington, DC.

Brown, K. A., Brick, P., & Blow, D. M. (1987) *Nature 326*, 416–418.

Bryan, P. N., Rollence, M.-L., Pantoliano, M. W., Wood, J., Finzel, B. C., Gilliland, G. L., Howard, A. J., & Poulos, T. L. (1987) *Proteins: Struct., Funct., Genet. 1*, 326–334.

Carter, P., & Wells, J. A. (1987) *Science 237*, 394.

Carter, P., & Wells, J. A. (1988) *Nature 332*, 564–568.

Carter, P., & Wells, J. A. (1990) *Proteins: Struct., Funct., Genet.* (in press).

Carter, P., Nilsson, B., Burnier, J. P., Burdick, D., & Wells, J. A. (1989) *Proteins: Struct., Funct., Genet. 6*, 240–248.

Carter, P. J., Winter, G., Wilkinson, A. J., & Fersht, A. R. (1984) *Cell 38*, 835–840.

Chothia, C., & Lesk, A. (1986) *EMBO J. 5*, 823–826.

Craik, C. S., Largman, C., Fletcher, T., Roczniak, S., Barr, P. J., Fletterick, R., & Rutter, W. J. (1985) *Science 228*, 291–297.

Cunningham, B. C., & Wells, J. A. (1987) *Protein Eng. 1*, 319–325.

Cunningham, B. C., Henner, D. J., & Wells, J. A. (1990) *Science 247*, 1461–1465.

Ebright, R. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 303–307.

Ebright, R. H., Kolb, A., Buc, H., Kunkel, T. A., Krakow, J. S., & Beckwith, J. (1987) *Proc. Natl. Acad. sci. U.S.A. 84*, 6083–6087.

Empie, M. W., & Laskowski, M., Jr. (1982) *Biochemistry 21*, 2274–2284.

Fersht, A. (1985) in *Enzyme Structure and Mechanism*, 2nd ed., Chapters 3, 12, and 13, W. H. Freeman and Co., New York.

Fersht, A. R. (1987) *Biochemistry 26*, 8031–8037.

Fersht, A. R., Wilkinson, A. J., Carter, P., & Winter, G. (1985) *Biochemistry 24*, 5858–5861.

Hecht, M. H., Sturtevant, J. M., & Sauer, R. T. (1986) *Proteins: Struct., Funct., Genet. 1* 43–46.

Horovitz, A., & Rigbi, M. (1985) *J. Theor. Biol. 116*, 149–159.

Howell, E. E., Villafranca, J. E., Warren, M. S., Oatley, S. J., & Kraut, J. (1986) *Science 231*, 1123–1128.

Hurle, M. R., Tweedy, N. B., & Matthews, C. R. (1986) *Biochemistry 25*, 6356–6360.

Imanaka, T., Shibazaki, M., & Takagi, M. (1986) *Nature 324*, 695–697.

Jencks, W. P. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 4046–4050.

Jones, M. D., Lowe, D. M., Borgford, T., & Fersht, A. R. (1986) *Biochemistry 25*, 1887–1891.

Katz, B. A. & Kossiakoff, A. (1986) *J. Biol. Chem. 261*, 15480–15485.

Laskowski, M., Jr., Tashiro, M., Empie, M. W., Park, S. J., Kato, I., Ardelt, W., & Mleczorek, M. (1983) in *Protease Inhibitors: Medical and Biological Aspects* (Katunuma, N., Ed.) pp 55–68, Japan Scientific Societies Press, Tokyo, Japan.

Laskowski, M., Jr., Kato, I., Ardelt, W., Cook, J., Denton, A., Empie, M. W., Kohr, W. J., Park, S. J., Parks, K., Schatzley, B. L., Schoenberger, O. L., Tashiro, M., Vichot, G., Whatley, H. E., Wieczorek, A., & Wieczorek, M. (1987) *Biochemistry 26*, 202–221.

Laskowski, M., Jr., Park, S. J., Tashiro, M., & Wynn, R. (1989) in *Protein Recognition of Immobilized Ligands*, UCLA Symposia on Molecular and Cellular Biology (Hutchens, T. W., Ed.) Vol. 80, pp 149–160, A. R. Liss, New York.

Leatherbarrow, R. J., Fersht, A. R., & Winter, G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7840–7844.

Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B., & Muller-Hill, B. (1990) *EMBO J. 9*, 615–621.

Liao, H., McKenzie, T., & Hageman, R. (1985) *Proc. Natl. Acad. Sci. U.S.A. 83*, 576–580.

Lowe, D. M., Fersht, A. R., Wilkinson, A. J., Carter, P., & Winter, G. (1985) *Biochemistry 24*, 5106–5109.

Matsumura, M., Yasumura, S., & Aiba, S. (1986) *Nature 323*, 356–358.

Matsumura, M., Signor, G., & Matthews, B. W. (1989) *Nature 342*, 291–294.

Matthews, B. W. (1987) *Biochemistry 26*, 6885–6888.

Mayer, R. J., Chen, J.-T., Taira, K., Fierke, C. A., & Benkovic, S. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7718–7720.

Mitchinson, C., & Baldwin, R. L. (1986) *Proteins: Struct., Funct., Genet. 1*, 23–33.

Nelson, H. C. M., & Sauer, R. T. (1985) *Cell 42*, 549–558.
Otlewski, J., & Laskowski, M., Jr. (1990) (submitted for publication).
Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L., & Bryan, P. N. (1989) *Biochemistry 28*, 7205–7213.
Russell, A. J., & Fersht, A. R. (1987) *Nature 328*, 496–500.
Sandberg, W. S., & Terwilliger, T. C. (1989) *Science 245*, 54–57.
Sarai, A., & Takeda, Y. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6513–6517.
Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.
Scrutton, N. S., Berry, A., & Perham, R. N. (1990) *Nature 343*, 38–43.
Shortle, D., & Meeker, A. K. (1986) *Proteins: Struct., Funct., Genet. 1*, 81–89.
Takeda, Y., Sarai, A., & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 439–443.
Wells, J. A., & Estell, D. A. (1988) *Trends Biochem. Sci. 13*, 291–297.
Wells, J. A., Cunningham, B. C., Graycar, T. P., & Estell, D. A. (1987a) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5167–5171.
Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., & Estell, D. A. (1987b) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1219–1223.
Wells, J. A., Cunningham, B. C., Graycar, T. P., Estell, D. A., & Carter, P. (1987c) *Cold Spring Harbor Symp. Quant. Biol. 52*, 647–652.
Wetzel, R., Perry, L. J., Baase, W. A., & Becktel, W. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 401–405.
Wilde, J. A., Bolton, P. H., Dell'Acqua, M., Hibler, D. W., Pourmotabbed, T., & Gerlt, J. A. (1988) *Biochemistry 27*, 4127–4132.
Wilkinson, A. J., Fersht, A. R., Blow, D. M., & Winter, G. (1983) *Biochemistry 22*, 3581–3586.
Wilkinson, A. J., Fersht, A. R., Blow, D. M., Carter, P., & Winter, G. (1984) *Nature 307*, 187–188.

Accelerated Publications

Role of Tyrosine M210 in the Initial Charge Separation of Reaction Centers of *Rhodobacter sphaeroides*†

Ulrich Finkele, Christoph Lauterwasser, and Wolfgang Zinth
Physik Department der Technischen Universität, D-8000 München 2, FRG
Kevin A. Gray and Dieter Oesterhelt*
Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG
Received May 22, 1990; Revised Manuscript Received July 12, 1990

ABSTRACT: Femtosecond spectroscopy was used in combination with site-directed mutagenesis to study the influence of tyrosine M210 (YM210) on the primary electron transfer in the reaction center of *Rhodobacter sphaeroides*. The exchange of YM210 to phenylalanine caused the time constant of primary electron transfer to increase from 3.5 ± 0.4 ps to 16 ± 6 ps while the exchange to leucine increased the time constant even more to 22 ± 8 ps. The results suggest that tyrosine M210 is important for the fast rate of the primary electron transfer.

The primary photochemical event during photosynthesis of bacteriochlorophyll- (Bchl-) containing organisms is a light-induced charge separation within a transmembrane protein complex called the reaction center (RC). The crystal structures of RC's from *Rhodopseudomonas* (*Rps.*) *viridis* and *Rhodobacter* (*Rb.*) *sphaeroides* have been solved to high resolution [reviewed in Deisenhofer and Michel (1989), Chang et al. (1986), Tiede et al. (1988), and Rees et al. (1989)]. The RC from *Rb. sphaeroides* contains three protein subunits referred to as L, M, and H, according to their respective mobilities in SDS–polyacrylamide gels. Associated with the L and M subunits are the cofactors, consisting of four Bchl *a*, two bacteriopheophytin (Bph) *a*, one atom of non-heme ferrous iron, two quinones (Q_A and Q_B), and in some species one carotenoid [reviewed in Parson (1987) and Feher et al. (1989)]. The cofactors are arranged in two branches (Figure 1) with an approximate C_2 axis of symmetry. The kinetic data support a model in which the primary electron transfer proceeds after light absorption by the primary donor [a special pair of Bchl referred to as P; reviewed in Kirmaier and Holten (1987)]. The absorption of light generates the excited electronic state P^*, which has a lifetime of approximately 3 ps. An electron is transferred from P along only one branch (the so-called A-branch). It is generally accepted that after approximately 3 ps the electron arrives at the Bph on the A-side (H_A) and after 220 ps it reaches Q_A. The role of the accessory Bchl located between P and H_A (referred to as B_A) has not been definitely assigned. Recently, we have shown that at room temperature an additional kinetic ($\tau \approx 0.9$ ps) component is detectable (Holzapfel et al., 1989). The spectral properties and the kinetic constants lead to the conclusion that the corresponding intermediate is the radical pair $P^+B_A^-$ (Holzapfel et al., 1990).

Additional intriguing points concerning the process of

† Financial support was from the Deutsche Forschungsgemeinschaft, SFB 143.
* To whom correspondence should be addressed.

0006-2960/90/0429-8517$02.50/0

14

Computational Complexity, Protein Structure Prediction, and the Levinthal Paradox

J. Thomas Ngo, Joe Marks, and Martin Karplus

1. Perspectives and Overview

A protein molecule is a covalent chain of amino acid residues. Although it is topologically linear, in physiological conditions it folds into a unique (though flexible) three-dimensional structure. This structure, which has been determined by x-ray crystallography and nuclear magnetic resonance for many proteins (Bernstein et al., 1977; Abola et al., 1987), is referred to as the native structure. As demonstrated by the experiments of Anfinsen and co-workers (Anfinsen et al., 1961; Anfinsen, 1973), at least some protein molecules, when denatured (unfolded) by disrupting conditions in their environment (such as acidity or high temperature) can spontaneously refold to their native structures when proper physiological conditions are restored. Thus, all of the information necessary to determine the native structure can be contained in the amino acid sequence.

From this observation, it is reasonable to suppose that the native fold of a protein can be predicted *computationally* using information only about its chemical composition. In particular, it should be possible to write down a mathematical problem that, when solved, gives the native conformation of the protein. This procedure would be self-contained, in the sense that no additional information about the biology of protein synthesis would be required. Further, it is reasonable to hope that this procedure could be accomplished without requiring an astronomical amount of computer resources, given the observation that polypeptide chains do fold to their

The Protein Folding Problem and Tertiary Structure Prediction
K. Merz, Jr. and S. Le Grand, Editors

Table 14-1. Glossary of problem-related terms.

problem instance The task of finding the global minimum[a] of a particular function $f(x_1, x_2, \ldots, x_n)$, e.g., the task of finding the value of x that will give the smallest possible value of $f(x) = 18x^7 + x^3 - 3x^2$. Note that the problem instance must be self-contained—everything necessary to specify the problem fully must appear.
problem A well-defined class of problem instances, e.g., the task of globally minimizing $f(x_1, x_2, \ldots, x_n)$, where f is any polynomial in the variables $x_1, x_2, \ldots, x_n$.
restriction Some stipulation about what constitutes an instance of the problem. In the example we have been using so far, it was stated that f is restricted to be a polynomial in one or more real variables.
restricted problem One problem, X, is said to be a restricted version of Y if X comprises a subset of the instances of Y. For example, the problem of globally minimizing quadratic functions is a restricted version of the problem of globally minimizing polynomial functions, because all quadratic functions are polynomials.
instance description The information needed to specify a particular instance of a given problem. In the example we have been using so far, the instance description might contain the following information: (a) the number of variables x_i is 1; (b) the polynomial is of degree 7; (c) the coefficients of the polynomial terms are, in descending order of degree, $\{18, 0, 0, 0, 1, -3, 0, 0\}$. Note that it is possible to construct a precise instance description like this only when the restrictions on a problem are well defined.
instance size The amount of memory needed to store a particular instance description.

[a] For simplicity in this discussion, we assume that the only computational tasks of interest to the reader are those that involve global optimization of a function.

native structures in an amount of time far less than required for an exhaustive search.

Based partly on these suppositions, computational chemists have developed and are refining expressions for the potential energy of a protein as a function of its conformation (Momany et al., 1975; Brooks et al., 1983; Weiner et al., 1986, for example). According to the Thermodynamic Hypothesis (Epstein et al., 1963), finding the global minimum of a protein's potential-energy function should be tantamount to identifying the protein's native fold.[1] (All notes are at the end of the chapter.) Unfortunately, the task of finding the global minimum of one of these functions is not easy, in

Table 14-2. Glossa

algorithm A and supplies an an algorithm is associated resu any sense to the
correctness *stance* if it giv instance. It is s instances of the it is understood
efficiency An to return *some* problem, within of the instance problem instan
It must be en we employ thr practical notio portional to n^1 practical; neve while it is usu time) algorithm problem instan algorithm is fa
guarantee A proved rigoro cient for a pro enough to be for all instanc
polynomially ed if there exi for every posi call a function functions that included in th is not polyno scription (suc and Johnson,
exhaustivene tively if it tes exhaustive wi be exhaustive

Table 14-2. Glossary of algorithm-related terms.

algorithm A computer program that takes an instance description as input and supplies an answer to the given problem instance as output. Note that an algorithm is always associated with a particular problem, including any associated restrictions; otherwise, the instance description may not make any sense to the program.

correctness An algorithm is said to be correct for a given problem *instance* if it gives the right answer when fed with the description of that instance. It is said to be correct for a *problem* if it is correct for all possible instances of the problem. When an algorithm is simply said to be *correct*, it is understood to be correct for the problem for which it was designed.

efficiency An algorithm is efficient for a given problem if is guaranteed to return *some* answer, right or wrong, to every possible instance of the problem, within an amount of time that is a polynomially bounded function of the instance size. Note that there is no notion of efficiency for a single problem instance.

It must be emphasized that this definition of the word "efficiency," which we employ throughout this review, does not necessarily correspond to the practical notion of efficiency. An algorithm whose running time is proportional to n^{100}, where n is the size of a problem instance, is far from practical; nevertheless, it is considered efficient by this definition. Thus, while it is usually reasonable to assume that an inefficient (exponential-time) algorithm is too slow to be of any practical use except for small problem instances, it is not always reasonable to assume that an efficient algorithm is fast enough to be of practical use.

guarantee A statement about the behavior of an algorithm that can be proved rigorously. For example, an algorithm is not considered to be efficient for a problem unless polynomial time bounds can be proved; it is not enough to be able to say that the algorithm has met the given time bounds for all instances of practical interest with which it has been tested.

polynomially bounded function A function $f(n)$ is polynomially bounded if there exists some polynomial function $g(n)$ such that $f(n) \leq g(n)$ for every positive value of n. In the interest of brevity, computer scientists call a function *polynomial* if it is polynomially bounded, even though some functions that are not normally considered polynomial (such as $\log n$) are included in this definition. Similarly, a function is called *exponential* if it is not polynomially bounded, even though some functions that fit this description (such as $n^{\log n}$) are not normally thought of as exponential (Garey and Johnson, 1979).

exhaustiveness An algorithm is said to search its solution space exhaustively if it tests every possible candidate solution. An algorithm that is exhaustive will always be correct; but to be correct, an algorithm need not be exhaustive.

particular because the potential-energy surface of a protein (as represented by an empirical potential) contains many local minima (Elber and Karplus, 1987, for example). Many clever and extremely creative techniques have been employed to try to escape from these local minima (Piela et al., 1989; Gordon and Somorjai, 1992; Head-Gordon and Stillinger, 1993), but so far no practical method for globally optimizing the potential energy of a protein has been produced.

The central question addressed in this review is this: Is there some clever algorithm, yet to be invented, that can find the global minimum of a protein's potential-energy function reliably and reasonably quickly? Or is there something intrinsic to the problem that prevents such a solution from existing?

Many measures of a problem's difficulty are possible. They range from the informal to the formal, and they focus on various sources of difficulty. Each can be useful in its own right. For example, proteins are certainly "hard to model with quantitative accuracy," the more realistic energy functions are "complicated to code as computer subroutines," and the algorithms that one uses to try to find the minima of these functions can consume seemingly unlimited amounts of supercomputer time.

Instead of such qualitative statements, the results reviewed here focus on a formal measure of difficulty called *intractability*. A problem is said to be *tractable* if there exists an algorithm for it that is *guaranteed* to be *correct* and *efficient*. It is said to be *intractable* if no such algorithm exists. (Several terms such as "correct" and "efficient" are used in a precise technical sense in this chapter. Tables 14-1 and 14-2 list their definitions.) When a problem is intractable, there is generally a rather unforgiving limit to the size of the problem instances that can be solved correctly before running times become astronomically large, and this limit is relatively uninfluenced by implementation details such as cleverly designed program codes and improvements in computer hardware.

It is widely believed that the problem of locating the global minimum of a protein's potential energy function is intractable by this definition. However, the conventional reasoning underlying this belief is fallacious. The conventional argument proceeds as follows. Although the bond lengths and angles in a protein can be predicted easily since they cannot vary much, torsion angles (rotations about bonds) are not easily predicted. Rotatable torsions tend to have three preferred values, none of which can be ruled out *a priori*. If there are N rotatable torsions in the protein molecule, then there are 3^N possible combinations of those torsions. To be sure to find the best combination, a computer program would have to try them all. But 3^N is a huge number for typical values of N because it is exponential in N. For

typical values of N ($\sim$ 100 or larger) and the speed of current hardware, the expected running time of this exhaustive algorithm is astronomically long.

This reasoning is fallacious. For some variants of the problem, it is wrong.[2] Every problem in combinatorial optimization[3] has an exponential number of candidate solutions. It will therefore require an exponential amount of time to solve *any* such problem by an exhaustive search of its candidate solutions. However, it is rarely necessary to proceed by such brute-force tactics (see Figure 14-6). With some thought, it is nearly always possible to do many orders of magnitude better than exhaustive search. Moreover, with many combinatorial optimization problems, algorithms can be found that are both efficient and correct (have polynomial time bounds and always give a right answer—see Table 14-2).

Is global potential-energy minimization *inherently impossible* to accomplish efficiently without sacrificing correctness, or is an efficient, correct algorithm waiting to be found? Recently, efforts have been made to answer this question using the formal tools of the theory of NP-completeness (Ngo and Marks, 1992; Unger and Moult, 1993). Introduced in the 1970's, NP-completeness theory (Lewis and Papadimitriou, 1978; Garey and Johnson, 1979; Lewis and Papadimitriou, 1981) was developed to help discover why certain problems in combinatorial optimization seem to be intractable, whereas others are not. A problem that is found to be NP-complete or NP-hard by an analysis of this type is intractable if P$\neq$NP. (The meaning of the proposition "P$\neq$NP" is summarized in Figure 14-1.)

One motivation for undertaking this line of inquiry pertains to the development of structure-prediction algorithms. Practical solutions to NP-complete problems do exist. They are compromises that entail well-understood tradeoffs between guarantees of efficiency, correctness, and generality (Papadimitriou and Steiglitz, 1982), and the forms of these compromises have been studied extensively. Thus, the mere fact that a problem is known to be NP-complete can guide algorithm developers to existing classes of heuristic solutions. Moreover, the *details* of an NP-completeness proof can expose the sources of a particular problem's complexity. With this knowledge, the algorithm developer can know in advance that certain algorithms are bound to fail, and might identify restricted forms of the problem that can be solved efficiently. These issues are discussed in Section 4.

The other goal is to obtain an improved understanding of the Levinthal paradox (Levinthal, 1968, 1969). The Levinthal paradox refers to the observation that although a protein is expected to require exponential time to achieve its native state from an arbitrary starting configuration, the process of folding is not observed to require exponential time. But why is a protein

expected to require exponential time to fold? The conventional justification for this premise requires the use of a model of protein behavior that leads to incorrect physical consequences. A reformulation of the Levinthal paradox with a more rigorous reason to expect exponential-time folding is discussed in Section 5.

"P≠NP"—What does it mean?

The proposition P≠NP, whose truth (or falsehood) has not been proved, is a pivotal conjecture in the theory of NP-completeness (Lewis and Papadimitriou, 1978). P is the class of problems for which correct algorithms with polynomial time bounds exist; NP is the class of problems for which a correct answer can be verified in polynomial time. The underlying theory and the precise meaning of the designation "NP" are too subtle to treat properly in this review, but the consequences of the proposition are easily summarized.

- Computer scientists have identified several classes of problems whose intractability is likely, but not yet proved. Two of the most important such classes are called NP-complete and NP-hard. Membership in these classes is well-defined whether or not P=NP. Hundreds of NP-complete and NP-hard problems have been identified; many of these problems are of great practical significance.
- If P=NP, then all NP-complete problems are efficiently solvable. If P≠NP, then all NP-complete problems are intractable.
- The class of NP-hard problems includes all NP-complete problems and others. An NP-hard problem can be thought of as being "at least as hard as" an NP-complete problem. If the NP-complete problems are intractable, then all of the NP-hard problems are intractable. But if the NP-complete problems are efficiently solvable, some NP-hard problems will still be intractable.
- (Almost) nobody believes that P=NP.

Figure 14-1. Meaning of the proposition "P≠NP."

The conclusions that may be drawn from the results described here are rigorous but qualified. In addition to reviewing the results themselves, much of the space in this chapter is devoted to identifying the caveats associated with each possible inference. There are some fundamental limitations of the scope of this approach that we state in advance.

First, the theory of NP-completeness can be used to address only certain aspects of the protein-folding problem. The protein-folding problem can be defined as encompassing, but not being limited to, the following questions:

1. Why does a
   small portion
   lated under p
2. What can be
   proteins reac
3. What accoun
4. Can a protein
   acid sequenc

The results that w
ture prediction (4
but they have lit
(1) and the pathv
that the four que
protein-folding p
that describes the
tive structure is
aspects of the pr
edited by Creigh

Second, it is i
lem has been shc
man, 1988) to pr
potential-energy
the solution in a
prediction, wher
is one approach
completeness th
prediction, the a
structure predict
algorithms in m
from the inadec
in reality, not fr
arise in the cor
formal bearing
self-contained[4]

It is also in
for a protein ha
is much closer
proof can, at t
than protein-stu
possibly more

1. Why does a protein have a unique native structure, i.e., why is such a small portion of a protein's conformational space significantly populated under physiological conditions?
2. What can be said about the pathway(s) in conformation space by which proteins reach their native states?
3. What accounts for the observed rate at which proteins fold?
4. Can a protein's three-dimensional structure be predicted from its amino acid sequence, and if so, how?

The results that we review here are related directly to questions about structure prediction (4) and indirectly to the consideration of folding rates (3), but they have little to do with the existence of unique native structures (1) and the pathway(s) by which a protein folds (2). (This is not to say that the four questions are unrelated to each other. All aspects of the protein-folding problem are determined by the potential-energy function that describes the polypeptide chain. Thus, the population of a unique native structure is clearly related to the prediction of that structure.) Many aspects of the protein-folding problem are addressed in a recent volume edited by Creighton (1992).

Second, it is incorrect to state that "the protein-structure prediction problem has been shown to be NP-hard." There are numerous approaches (Fasman, 1988) to protein-structure prediction that either do not employ global potential-energy minimization at all, or include stipulations on the nature of the solution in addition to the energy function itself. Secondary-structure prediction, when based on statistical rules derived from known structures, is one approach in which no potential-energy function is used. While NP-completeness theory certainly could be used to analyze secondary-structure prediction, the analysis might be irrelevant because models for secondary-structure prediction are usually designed with efficient (usually linear-time) algorithms in mind. The difficulty of predicting secondary structure arises from the inadequacy of the underlying models in predicting what occurs in reality, not from the time required to solve the computational tasks that arise in the context of those models. The results reviewed here have a formal bearing only on algorithms that operate by attempting to solve a self-contained[4] global minimization problem.

It is also incorrect to state that "global potential-energy minimization for a protein has been shown to be NP-hard," even though this statement is much closer to the truth than the previous one. An NP-completeness proof can, at best, address a form of the problem that is more general than protein-structure prediction by energy minimization—and therefore, possibly more difficult. Put another way, protein-structure prediction might

be an easy restricted form (special case) of the problems (Sections 3.3, 3.4, and 3.5) that are known to be NP-hard. This subtle point, which we discuss in Sections 4 and 6, turns out to be central to understanding the limitations of the results.

Third, for some purposes, intractability itself is not as bad as it sounds. The intractability of a problem means that no algorithm for it can be efficient and correct. These qualities would make a protein-structure prediction algorithm, perhaps literally, "too good to be true." Most developers of protein-structure prediction algorithms gave up on such high standards long ago; they focus efforts on developing algorithms that fall short of the ideal in some way. Thus, the NP-hardness of a problem is a somewhat weak statement, even given the very likely assumption that P≠NP. An objective of this review is to explore what may be inferred from this weak statement, given that it is one of very few statements about the difficulty of protein-structure prediction that are known rigorously to be true.

In writing this review, we have tried to focus on objectives that are appropriate for this young line of inquiry. First, we believe that the question most important to the reader, given that he accepts that the proofs being reviewed are mathematically correct, is what they mean. Accordingly, we include an exposition of the theory of NP-completeness (Section 2) that is intended to give an accurate picture of the form of an NP-completeness proof without becoming mired in technical details. Similarly, in explaining the proofs themselves (Sections 3.3, 3.4, and 3.5), we describe the essential steps—it is impossible to understand exactly what is proved otherwise—but do not attempt to persuade the reader of their mathematical correctness. We invite the reader who is interested in the technical details to refer to the original papers (Ngo and Marks, 1992; Fraenkel, 1993; Unger and Moult, 1993) and to existing references on the theory of NP-completeness (Lewis and Papadimitriou, 1978; Garey and Johnson, 1979; Lewis and Papadimitriou, 1981).

Second, we believe that one of the functions of a review is to reinterpret and evaluate existing results. Therefore, interspersed with a straightforward recitation of the facts, the reader will find our opinions and speculations about the implications of this line of reasoning, both for the development of algorithms (Section 4) and for the behavior of real proteins (Section 5).

Third, from the limitations of the existing results, it is clear that the use of computational complexity theory for tasks in protein-structure prediction is by no means a closed book. In Section 6 we point out areas in which continued analysis might be of value, particularly given the results that have already been established.

which depends steeply on the energy gap U. Given the assumptions that $N = 100$ and $n = 2$, it was found that in the limit $U \to 0$, the first-passage time is nearly 10^{30} years. However, a modest change to the value of U, say $U = 2kT$, lowers the first-passage time to under one second. (The base of the exponential, $1 + n\exp(-U/kT)$, is equal to 3 when $U = 0$, but 1.27 when $U = 2kT$.)

The analysis of Zwanzig et al. resolves a form of the Levinthal paradox in which the absence of clues about the form of the native state is the sole basis for expecting exponential-time folding. However, it does not resolve the form of the paradox based on computational complexity, since the optimization problem implied by the underlying model can be solved trivially in linear time. The reason for the tractability of the underlying model is the lack of long-range interactions, which are critical to rendering PSP NP-hard (Ngo and Marks, 1992), and essential for cooperativity (Karplus and Shakhnovich, 1992).

6. Future Work

It is not known whether there exists an efficient algorithm for predicting the structure of a given protein from its amino acid sequence alone. Decades of research have failed to produce such an algorithm, yet Nature seems to solve the problem. Proteins do fold! The "direct" approach to structure prediction, that of directly simulating the folding process, is not yet possible because contemporary hardware falls eight to nine orders of magnitude short of the task. However, while this difference is large, it is not astronomical. Would this "direct" approach constitute an efficient and correct algorithm for protein-structure prediction? Too little is known about protein folding, and about the future of computing technology, to be able to answer this question at this time.

The results reviewed here (Section 3) do not completely rule out the existence of a protein-structure prediction algorithm that is both efficient and correct, in the precise senses of those words used throughout this chapter. In particular, it remains formally possible that there is a restricted form of PSP that is efficiently solvable, but subsumes protein-structure prediction. How can this possibility be investigated?

A standard strategy in the analysis of any NP-hard problem is to examine restricted forms of the problem systematically, classifying each as tractable or NP-hard, and thereby exposing the sources of the complexity. Barahona's results with Ising spin-glass models, which were described briefly in Section 4, are exemplary of this approach. While the particular

restrictions chosen by Barahona for spin glasses (reduction of dimensionality and removal of the magnetic field) are not suitable for protein-structure prediction, the overall strategy of examining restricted forms is appropriate. Some restricted form of PSP in which compactness plays a critical role is a candidate for this type of analysis (Section 4.6).

The approach of considering restricted forms has worked well for dozens of important problems that are relatively "clean" and abstract (Garey and Johnson, 1979), but it may be difficult to pursue in the case of protein-structure prediction. In the former case, the problem shown to be NP-hard is usually as general as would actually be required in practice. In the latter case, what is desired is not an algorithm that can handle all possible instances of PSP (Section 3), but merely one that works for proteins. Thus, the fact that PSP is a generalization of protein-structure prediction makes the result that PSP is NP-hard less limiting than it could be.

Ideally, one would like to demonstrate the NP-hardness of a problem that is more *specific*, not more general, than protein-structure prediction, because that would automatically prove the NP-hardness of protein-structure prediction itself. This would entail finding an efficient transformation from some existing NP-complete problem that generates instances of PSP that are proteins by every conceivable criterion.[38] It is difficult to see how such a transformation might proceed.[39]

An alternative approach that may be nearly as instructive is to use the currently available result regarding PSP as a baseline in a continuing comparative analysis—to find restricted forms of PSP that are NP-hard but as specialized as possible, and to find others that are tractable but as general as possible. The motivations for pursuing this methodology are both practical and theoretical:

- Every NP-hardness result permits us to know in advance that a certain group of algorithms is likely to fail, and is therefore not worth pursuing (Section 4).
- Conversely, every NP-hardness result helps identify a source of complexity in protein-structure prediction, and therefore what must be stripped away from the problem before it is reasonable to attempt efficient solution.

  The work of Finkelstein and Reva (1992) is a good example; an approach to structure prediction with a guaranteed polynomial time bound was developed. The critical assumption behind the algorithm is that only nonbonded interactions between nearest neighbors along the chain are significant. Because of this assumption, the algorithm cannot solve all instances of PSP, but instead is restricted to instances in

which only nonbonded interactions between nearest neighbors along the chain are nonzero.[40] This violates the requirements of the reduction from Partition to PSP, in which nonbonded interactions between sites distant from each other along the chain are essential. Thus, the problem is similar in character to that examined by Zwanzig et al. (Section 5.3). While the Finkelstein–Reva algorithm was not inspired by an NP-hardness result, the underlying strategy is similar to how NP-hardness results might be used; they removed from the problem what they observed to be a source of complexity. However, in this case, removing the source of complexity led to a problem different from that posed by protein folding, in which long-range interactions play an essential role.

- The NP-hardness of PSP serves as the premise for a reformulation of the Levinthal paradox (Section 5), whose conventional form is based on a model of folding that is in conflict with known experimental results. A motivation for pursuing an analysis of the computational complexity of protein-structure prediction is to assist in the constructive role of the Levinthal paradox—to help focus attention on the key questions in protein folding.

  A small number of reasonably well-defined potential resolutions to the computational-complexity form of the Levinthal paradox were listed in Section 5. One of the possible resolutions is that protein-structure prediction is tractable. NP-hardness results with restricted forms of PSP would make that possible resolution less likely, thus lending credence to the alternatives.

  Attempts to resolve the Levinthal paradox, which play a valid and useful role in helping to understand how proteins fold, can lead to confusion because the premises of the original form of the paradox are not well formulated. In particular, one such proposed resolution (Zwanzig et al., 1992) can be shown unequivocally not to resolve the computational complexity form of the paradox, and in related arguments (Karplus and Shakhnovich, 1992) has been shown to lead to physically incorrect consequences (Section 5.3). For the paradox to be meaningful, it must be "falsifiable"—it must be possible to know when the paradox has been resolved.

In addition to restricted forms of PSP, it would be useful to know the computational complexity of other tasks in structure prediction that appear easier than the general problem, but whose complexities are none the less uncertain.

The task of cor
of a protein's backt
ing simulated anne
ing effects may sui
in a protein's core
lem is unknown.
of possible side-ch
sparse, and consis
instance, with Isin
ity (Garey and Joh
and Wagner, 1991
borhood interactic
other hand, many
are tractable if res
and Johnson, 197
ceptable set of si
(One currently kı
based on backbor
angles [Dunbrack
complexity of sic
oper in the quan
necessary, given

Acknowledgment
and providing a
kindly provided a
Shakhnovich, an
JTN is grateful f
Foundation. Thi
Science Foundat

NOTES

[1] The Thermod
figuration of glob
a protein's states
use of potential e
potential energies

[2] For example,
the chain are sign
(Finkelstein and I

[3] The term *com*
the solution spac

The task of computing side-chain conformations given full knowledge of a protein's backbone conformation is one such problem. Case studies using simulated annealing (Lee and Subbiah, 1991) have suggested that packing effects may suffice to determine, in part, the side-chain conformations in a protein's core. The computational complexity of this packing problem is unknown. Because only short-range effects are present, the graph of possible side-chain–side-chain interactions can be known in advance, is sparse, and consists of vertices of low degree. Previous experience—for instance, with Ising spin-glass models (Barahona, 1982), graph colorability (Garey and Johnson, 1979, p. 191) and cartographic labeling (Formann and Wagner, 1991; Marks and Shieber, 1991)—illustrates that such neighborhood interactions can, on their own, give rise to NP-hardness. On the other hand, many problems that contain such neighborhood interactions are tractable if restrictions can be placed on the nature of the graph (Garey and Johnson, 1979), suggesting that the problem of finding a mutually acceptable set of side-chain conformations for a protein could be tractable. (One currently known algorithm for predicting side-chain conformations based on backbone positions achieves 70% to 80% accuracy for χ_1 and χ_2 angles [Dunbrack and Karplus, 1993].) Not knowing the computational complexity of side-chain structure prediction leaves the algorithm developer in the quandary of not knowing whether inexact methods are truly necessary, given the possible existence of a superior exact algorithm.

Acknowledgments. We thank Ron Unger for answering detailed questions and providing a preprint (Unger and Moult, 1993). Aviezri Fraenkel also kindly provided a preprint (Fraenkel, 1993). We thank Harry Lewis, Eugene Shakhnovich, and Jim Clark for reading and commenting on the manuscript. JTN is grateful for a Graduate Fellowship from the Fannie and John Hertz Foundation. This research was supported in part by grants from the National Science Foundation and the National Institutes of Health.

NOTES

[1] The Thermodynamic Hypothesis states that a protein's native fold is the configuration of globally minimal free energy. However, it is generally assumed that a protein's states of lowest free energy are similar enough in entropy to justify the use of potential energies instead of free energies as a computational convenience; potential energies are much faster and more straightforward to compute.

[2] For example, if only nonbonded interactions between nearest neighbors along the chain are significant, the global minimum structure can be predicted efficiently (Finkelstein and Reva, 1992).

[3] The term *combinatorial optimization* is normally reserved for problems in which the solution space is discrete. Throughout this chapter we use the term to refer

CULTURE OF ANIMAL CELLS

A Manual of Basic Technique

R. Ian Freshney

Department of Clinical Oncology
Cancer Research Campaign Laboratories
University of Glasgow

Alan R. Liss, Inc., New York

Address all Inquiries to the Publisher
Alan R. Liss, Inc., 150 Fifth Avenue, New York, NY 10011

Printed in the United States of America

Second Printing, April 1984
Third Printing, November 1984
Fourth Printing, November 1986

Library of Congress Cataloging in Publication Data

Freshney, R. Ian.
Culture of animal cells.

Bibliography: p.
Includes index.
1. Tissue culture. 2. Cell culture. I. Title.
QH585.F74 1983 591'.07'24 82-24960
ISBN 0-8451-0223-0

Other areas of major interest include the study of cell interactions and intracellular control mechanisms in cell differentiation and development [Auerbach and Grobstein, 1958; Cox, 1974; Finbow and Pitts, 1981] and attempts to analyze nervous function [Bornstein and Murray, 1958; Minna et al., 1972]. Progress in neurological research has, however, not had the benefit of working with propagated cell lines as propagation of neurons has not so far been possible *in vitro* without resorting to the use of transformed cells (see Chapter 20).

Tissue culture technology has also been adopted into many routine applications in medicine and industry. Chromosomal analysis of cells derived from the womb by amniocentesis can reveal genetic disorders in the unborn child, viral infections may be assayed qualitatively and quantitatively on monolayers of appropriate host cells, and the toxic effects of pharmaceutical compounds and potential environmental pollutants can be measured in colony-forming assays.

Further developments in the application of tissue culture to medical problems may follow from the demonstration that cultures of epidermal cells form functionally differentiated sheets in culture [Green et al., 1979], and endothelial cells may form capillaries [Folkman and Haudenschild, 1980], suggesting possibilities in homografting and reconstructive surgery using an individual's own cells. The introduction of heterologous genetic material into mammalian cells [Willecke et al., 1979; Wigler et al., 1979], although somewhat overshadowed by current propagation in bacteria, may yet prove a desirable means for producing biologically significant compounds such as growth hormone and insulin. Similarly, the production of monoclonal antibodies [Kohler and Milstein, 1975] in hybrids between human plasma cells and human myeloma cells may prove a valuable technique for the production of specific antibodies.

It is clear that the study of cellular activity in tissue culture may have many advantages; but in summarizing these, below, considerable emphasis must also be placed on its limitations, in order to maintain some sense of perspective.

ADVANTAGES OF TISSUE CULTURE

Control of the Environment

The two major advantages, as implied above, are the control of the physicochemical environment (pH, temperature, osmotic pressure, O_2, CO_2 tension), which may be controlled very precisely, and the physiological conditions, which may be kept relatively constant but cannot always be defined. Most media still require supplementation with serum which is highly variable [Olmsted, 1967; Honn et al., 1975], and contains undefined elements such as hormones and other regulatory substances. Gradually, however, the functions of serum are being understood; and as a result, it is being replaced by defined constituents [Birch and Pirt, 1971; Ham and McKeehan, 1978; Barnes and Sato, 1980].

Characterization and Homogeneity of Sample

Tissue samples are invariably heterogeneous. Replicates even from one tissue vary in their constituent cell types. After one or two passages, cultured cell lines assume a homogeneous, or at least uniform, constitution as the cells are randomly mixed at each transfer and the selective pressure of the culture conditions tends to produce a homogeneous culture of the most vigorous cell type. Hence, at each subculture each replicate sample will be identical, and the characteristics of the line may be perpetuated over several generations. Since experimental replicates are virtually identical, the need for statistical analysis of variance is seldom required.

Economy

Cultures may be exposed directly to a reagent at a lower and defined concentration, and with direct access to the cell. Consequently, less is required than for injection *in vivo* where >90% is lost by excretion and distribution to tissues other than those under study.

DISADVANTAGES

Expertise

Culture techniques must be carried out under strict aseptic conditions, because animal cells grow much less rapidly than many of the common contaminants such as bacteria, molds, and yeasts. Furthermore, unlike microorganisms, cells from multicellular animals do not exist in isolation, and consequently, are not able to sustain independent existence without the provision of a complex environment, simulating blood plasma or interstitial fluid. This implies a level of skill and understanding to appreciate the requirements of the system and to diagnose problems as they arise. Tissue culture should not be undertaken casually to run one or two experiments.

Quantity

A major limitation of cell culture is the expenditure of effort and materials that goes into the production of relatively little tissue. A realistic maximum per batch for most small laboratories (2 or 3 people doing tissue culture) might be 1–10 g of cells. With a little more effort and the facilities of a larger laboratory, 10–100 g is possible; above 100 g implies industrial pilot plant scale, beyond the reach of most laboratories, but not impossible if special facilities are provided.

The cost of producing cells in culture is about ten times that of using animal tissue. Consequently, if large amounts of tissue (>10 g) are required, the reasons for providing them by tissue culture must be very compelling. For smaller amounts of tissue ($\leqslant 10$ g), the costs are more readily absorbed into routine expenditure; but it is always worth considering whether assays or preparative procedures can be scaled down. Semimicro- or micro-scale assays can often be quicker due to reduced manipulation times, volumes, centrifuge times, etc. and are often more readily automated (see under Microtitration, Chapter 19).

Instability

This is a major problem with many continuous cell lines resulting from their unstable aneuploid chromosomal constitution. Even with short-term cultures, although they may be genetically stable, the heterogeneity of the cell population, with regard to cell growth rate, can produce variability from one passage to the next. This will be dealt with in more detail in Chapters 12 and 18.

MAJOR DIFFERENCES *IN VITRO*

Many of the differences in cell behavior between cultured cells and their counterparts *in vivo* stem from the dissociation of cells from a three-dimensional geometry and their propagation on a two-dimensional substrate. Specific cell interactions characteristic of the histology of the tissue are lost, and, as the cells spread out, become mobile and, in many cases, start to proliferate, the growth fraction of the cell population increases. When a cell line forms it may represent only one or two cell types and many heterotypic interactions are lost.

The culture environment also lacks the several systemic components involved in homeostatic regulation *in vivo*, principally those of the nervous and endocrine systems. Without this control, cellular metabolism may be more constant *in vitro* than *in vivo*, but may not be truly representative of the tissue from which the cells were derived. Recognition of this fact has led to the inclusion of a number of different hormones in culture media (see Chapter 9) and it seems likely that this trend will continue.

Energy metabolism *in vitro* occurs largely by glycolysis, and although the citric acid cycle is still functional it plays a lesser role.

It is not difficult to find many more differences between the environmental conditions of a cell *in vitro* and *in vivo* and this has often led to tissue culture being regarded in a rather skeptical light. Although the existence of such differences cannot be denied, it must be emphasized that many specialized functions are expressed in culture and as long as the limits of the model are appreciated, it can become a very valuable tool.

Origin of Cells

If differentiated properties are lost, for whatever reason, it is difficult to relate the cultured cells to functional cells in the tissue from which they were derived. Stable markers are required for characterization (see Chapter 15); and in addition, the culture conditions may need to be modified so that these markers are expressed (see next chapter).

DEFINITIONS

There are three main methods of initiating a culture [Schaeffer, 1979] (see Glossary and Fig. 1.2): (1) *Organ culture* implies that the architecture characteristic of the tissue *in vivo* is retained, at least in part, in the culture. Toward this end, the tissue is cultured at the liquid/gas interface (on a raft, grid, or gel) which favors retention of a spherical or three-dimensional shape. (2) In *primary explant culture* a fragment of tissue is placed at a glass (or plastic)/liquid interface where, following attachment, migration is promoted in the plane of the solid substrate. (3) *Cell culture* implies that the tissue or outgrowth from the primary explant is dispersed (mechanically or enzymatically) into a cell suspension which may then be cultured as an adherent monolayer on a solid substrate, or as a suspension in the culture medium.

Organ cultures, because of the retention of cell interactions as found in the tissue from which the culture was derived, tend to retain the differentiated properties of that tissue. They do not grow rapidly (cell proliferation is limited to the periphery of the explant and is restricted mainly to embryonic tissue) and hence cannot

IN

Alexandria, VA 22313-1450
If Undeliverable Return in Ten Days

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

AN EQUAL OPPORTUNITY EMPLOYER

UNITED STATES POSTAGE U.S. OFFICIAL PENALTY F PRIVATE USE PITNEY BOWES
02 1A $ 02.3
0004204478 SEP 22
MAILED FROM ZIP CODE 22

RETURN TO SENDER
UNDELIVERABLE AS ADDRESSED